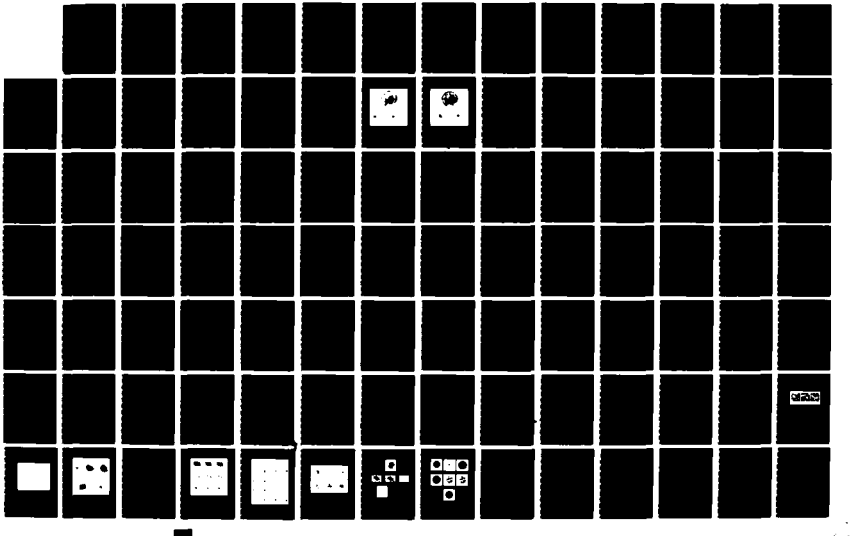
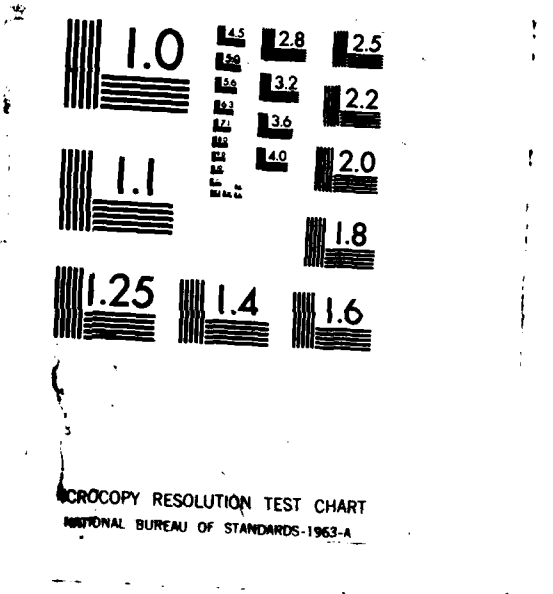
AD-A174 100 DIFFRACTION-LIMITED IMAGING OF SPACE OBJECTS III(U) 1/3
ENVIRONMENTAL RESEARCH INST OF MICHIGAN ANN ARBOR
J R FIENUP ET AL OCT 86 ERIM-161900-20-F
UNCLASSIFIED AFOSR-TR-86-2109 F49620-82-K-0018 F/G 14/5 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
7.1
CROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

(2)

Final Scientific Report

DIFFRACTION-LIMITED IMAGING OF SPACE OBJECTS III

1 March 1982 through 31 October 1985

J.R. FIENUP AND C.C. WACKERMAN
Advanced Concepts Division

OCTOBER 1986

Director, Physical and Geophysical Sciences
Air Force Office of Scientific Research/NP
Building 410, Bolling AFB, D.C. 20332
Contract No. F49620-82-K-0018

ENVIRONMENTAL
RESEARCH INSTITUTE OF MICHIGAN
BOX 8618 • ANN ARBOR • MICHIGAN 48107

SECURITY CLASSIFICATION OF THIS PAGE

AD-A174 100

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) 161900-20-F	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Environmental Research Institute of Michigan 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research/NP
6c. ADDRESS (City, State and ZIP Code) P.O. Box 8618 Ann Arbor, MI 48107	7b. ADDRESS (City, State and ZIP Code) Building 410 Bolling AFB, DC 20332
8a. NAME OF FUNDING/SPONSORING ORGANIZATION 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-82-K-0018
8c. ADDRESS (City, State and ZIP Code)	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.

11. TITLE (Include Security Classification)
Diffraction-Limited Imaging of Space Objects III

12. PERSONAL AUTHOR(S)
Fienup, James R. and Wackerman, Christopher C.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Final Scientific Report	FROM Mar 82 TO Oct 85	1986, October	221 + vii

16. SUPPLEMENTARY NOTATION
The AFOSR project scientist was Dr. Henry Radoski

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Phase retrieval, Space object imaging, Reconstruction algorithms, Image reconstruction, Astronomical speckle interferometry, Atmospheric Turbulence
20	06		
03			

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

This report describes the results of a three-year research program to investigate methods for obtaining diffraction-limited images of space objects, despite the turbulent atmosphere, by reconstructing images from data provided by optical interferometers (particularly stellar speckle interferometry). Major accomplishments include the following: (1) A new closed-form recursive algorithm was invented for reconstructing sampled objects having latent reference points. (2) Improvements in the iterative Fourier transform algorithm were devised, solving the stagnation problems of stripes and of simultaneous twin images. (3) Uniqueness of the reconstructed image was demonstrated empirically. (4) Reconstruction of complex-valued objects was shown to be possible. (5) The Hayes-Quatieri recursive algorithm was shown to suffer from a uniqueness problem, and that algorithm was generalized. (6) An improved method for estimating the object's Fourier modulus from stellar speckle interferometry data was devised. In addition, other reconstruction algorithms were investigated.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL	22b. TELEPHONE NUMBER (Include Area Code)	22c. OFFICE SYMBOL
Dr. Henry Radoski	(202) 767-4906	NP

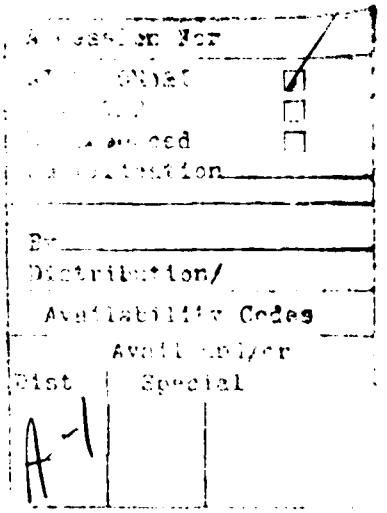
DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

FOREWORD

This report was prepared by the Optical Science Laboratory of the Advanced Concepts Division of the Environmental Research Institute of Michigan. The work was sponsored by the Air Force Office of Scientific Research/AFSC, United States Air Force, under Contract No. F49620-82-K-0018.

This final scientific report covers work performed between 1 March 1982 and 31 October 1985. The contract monitor is Dr. Henry Radoski, Directorate of Physical and Geophysical Sciences, AFOSR/NP, Building 410, Bolling Air Force Base, D.C. 20332. The principal investigator is James R. Fienup. Major contributors to the effort are James R. Fienup and Christopher C. Wackerman. Additional contributors to the effort are Thomas R. Crimmins, John D. Gorman, Susan C. Elm, William Licata and Raja Roumaya.



CONTENTS

LIST OF FIGURES

SUMMARY

This report describes the results of a three-year research program to investigate methods for obtaining diffraction-limited images of space objects, despite the turbulent atmosphere, by reconstructing images from data provided by optical interferometers (particularly stellar speckle interferometry). Major accomplishments include the following:

1. A new closed-form recursive algorithm was invented for reconstructing sampled objects having latent reference points.
2. Improvements in the iterative Fourier transform algorithm were devised, solving the stagnation problems of stripes and of simultaneous twin images.
3. Uniqueness of the reconstructed image was demonstrated empirically.
4. Reconstruction of complex-valued objects was shown to be possible.
5. The Hayes-Quatieri recursive algorithm was shown to suffer from a uniqueness problem, and that algorithm was generalized.
6. An improved method for estimating the object's Fourier modulus from stellar speckle interferometry data was devised.

In addition, other reconstruction algorithms were investigated.

DIFFRACTION-LIMITED IMAGING OF SPACE OBJECTS
1 March 1982 to 31 October 1985

I
INTRODUCTION AND OBJECTIVES

This report describes the results of a three-year research program to investigate methods for obtaining diffraction-limited images of space objects, despite the turbulent atmosphere, by reconstructing images from data provided by optical interferometers (particularly stellar speckle interferometry).

Atmospheric turbulence typically limits the angular resolution of earth-bound optical telescopes to one second of arc or worse, which is fifty times poorer than the theoretical diffraction limit of a 5-meter optical telescope. It is possible to gather diffraction-limited information through the turbulent atmosphere by a variety of interferometric techniques, including Michelson stellar interferometry [1], intensity interferometry [2], amplitude interferometry [3], and stellar speckle interferometry [4, 5]. However, this diffraction-limited information is in the form of the modulus (magnitude) of the Fourier transform of the object being viewed. Until recently only the autocorrelation of the object, but not the object itself, could be reconstructed from this data, except for special cases.

In recent years, an iterative method [6-8] has been developed for reconstructing an object from its Fourier modulus, thereby making possible the reconstruction of diffraction-limited imagery from interferometer data. The algorithm utilizes the measured Fourier modulus data as well as (1) the a priori information that the object's spatial (or angular) brightness distribution is a non-negative function and (2) information about the object's diameter which can be computed from the autocorrelation function. A detailed

description of the problem of image reconstruction from earth-bound optical telescopes including the iterative Fourier transform algorithms and other algorithms is given in Appendix A [9]. Further details on the iterative algorithm and its numerous applications were described in detail in Appendix A of last year's interim scientific report [10, 11].

The goal of the program is to further investigate and develop this method of obtaining diffraction-limited images. Included in the three-year program are investigations into improving the reconstruction algorithm, developing methods for processing astronomical data, and studying the uniqueness of the reconstruction. The first two years effort were described in the interim technical reports [12, 10]. In the third year of the effort, the emphasis was on developing improved reconstruction algorithms, demonstrating uniqueness empirically and processing atmospherically degraded data. This report summarizes the results of the entire three-year program.

The research accomplishments are summarized in Section 2 and are described in more detail in the Appendices. References are listed at the end of the report.

In overview, the results of the research program provide significant advancement in the understanding of the reconstruction problem and in the techniques available to solve it, which contribute toward a capability of reconstructing diffraction-limited images despite the presence of atmospheric turbulence. The results also serve as an important vehicle for further applications, such as the reconstruction of diffraction-limited images from degraded data provided by an active imaging system that suffers from phase errors or even total loss of phase information.

II
RESEARCH ACCOMPLISHMENTS

The major research results accomplished during the three-year program are as follows.

1. A new closed-form recursive algorithm was invented for reconstructing sampled objects having latent reference points [13].
2. Improvements in the iterative Fourier transform algorithm were devised, solving the stagnation problems of stripes and of simultaneous twin images [14].
3. Uniqueness of the reconstructed image was demonstrated empirically for a class of satellite objects [15].
4. Reconstruction of complex-valued objects was shown to be possible [16].
5. The Hayes-Quatieri recursive reconstruction algorithm, which assumes knowledge of the values of the boundary of the object, was shown to suffer from a uniqueness problem, and that algorithm was generalized [17].
6. An improved method for estimating the object's Fourier modulus from stellar speckle interferometry data was developed [18].

Additional research accomplishments are as follows.

7. A new version of the iterative Fourier transform algorithm, the relaxation algorithm, was developed.
8. Work progressed toward understanding the uniqueness properties of the reconstructed image in the 1-D case [19].
9. Alternative iterative algorithms were devised.

10. The shift-and-add algorithm was implemented and tested on a complicated extended object.

11. Astronomical data taken with the PAPA detector was analyzed.

Publications arising from this work are References 9 and 13-30. Reference 9 and 13 through 18 are included as Appendices A through G, respectively.

Reference 26 was an invited paper to the ICO-13 meeting in Sapporo, Japan in 1984. At the meeting, Dr. James R. Fienup received the International Commission for Optics' 1983 International Prize in Optics for his work in phase retrieval, which was largely the work performed on this and the previous associated AFOSR contract [32].

The eleven topics mentioned above are briefly described in the remainder of this section and are described in more detail in the Appendices and in the Interim Reports [10, 12].

2.1 NEW RECURSIVE ALGORITHM

A new recursive algorithm has been developed which is capable of reconstructing in closed form an object from its autocorrelation function, which can be computed from the modulus of its Fourier transform. It works for objects having latent reference points -- unresolved points within the object field that are not sufficiently far from the main part of the object to satisfy the condition for holography, but satisfying weaker conditions. A more detailed description of the algorithm is given in Appendix B.

The recursive algorithm was coded on a computer and exercised on two different types of objects using autocorrelations having a variety of signal-to-noise ratios. As expected, the recursive algorithm was highly sensitive to noise. As shown in Section 3 of [10], the RMS error of the reconstructed image was 0.0400, 0.1390 and 0.6088 when the RMS error of the data was 0.005175, 0.01795 and 0.05585,

respectively. This makes the recursive algorithm impractical compared with the more robust iterative transform algorithm. Nevertheless, the recursive algorithm is useful in showing that some objects are uniquely related to their Fourier modulus.

2.2 IMPROVEMENTS IN THE ITERATIVE ALGORITHM

In some cases, the iterative transform algorithm stagnates without reaching a solution. That is, on successive iterations the output image changes very little even though it is not a solution, i.e., it does not agree simultaneously with the Fourier modulus data and the object-domain constraints. Two modes of stagnation were identified, and methods were developed to overcome both of them.

When the object has symmetric support (the support is the set of points over which the object is nonzero), the iterative algorithm often gets into a stagnation mode for which the output image has features of both the object and its twin image (i.e., the object rotated by 180°). This stagnation problem of simultaneous twin images can be overcome by using, for a few iterations, a reduced-area support constraint that breaks the symmetry of the stagnation mode. Use of the defogging method of Bates and Fright [33], as described in Section 4 of [10], also helps. Since few space objects have precisely symmetric supports, this mode of stagnation does not usually cause a problem when imaging space objects; it does occur frequently for made-up problems such as a photographic image bounded by a rectangle.

Another mode of stagnation is characterized by an output image that resembles the true object but having a pattern of stripes superimposed. This problem occurs frequently for imaging satellites. In most cases the stripes are of such low contrast as to be hardly noticeable, and so they are not objectionable. In some cases, however, the stripes are of sufficiently high contrast to significantly

degrade the quality of the image. After a few false starts, two methods of overcoming this stagnation problem were developed. In the "voting method," three output images having different sets of stripes are produced by running the iterative transform algorithm three times each with a different starting input of random numbers. At each point in the Fourier domain, the complex Fourier transforms of the three striped images are compared. The value that is farthest from the other two is discarded and the other two are averaged to arrive at a composite Fourier transform. In the "patching method" only two different striped images are employed. The Fourier transform of the area outside the object's support is used to identify the regions in the Fourier domain in which the phase is in error (which causes the stripes). A composite Fourier transform is formed by patching together the "good" regions of the two Fourier transforms. In addition, the origins of the stripes are better understood. These results are described in detail in Appendix C [14].

2.3 UNIQUENESS DEMONSTRATION

In an earlier effort [34] an attempt was made to demonstrate empirically the uniqueness of imagery reconstructed from Fourier modulus data by performing reconstruction experiments and seeing whether the reconstructed images resembled the original object or a (possible) alternative solution. Although those results looked promising, they were hampered by the stripes stagnation problem discussed above. With the improved reconstruction algorithm in hand (see Appendix C), those uniqueness experiments were redone. The results, as shown in Appendix D [15], demonstrate that for satellite imagery viewed passively, the reconstructed image is usually unique in a practical sense.

2.4 RECONSTRUCTION OF COMPLEX-VALUED OBJECTS

The vast majority of the research effort was involved with reconstructing real, nonnegative images as would be the case for passive incoherent imaging. There are other remote sensing applications, however, for which the object to be reconstructed is inherently complex-valued. These are active, coherent imaging applications such as synthetic aperture radar (SAR), imaging a laser-illuminated object, and wavefront sensing. Despite the loss of the powerful nonnegativity constraint, it was found that if one had a support constraint that was sufficiently tight and of a strong type, then it was possible to reconstruct a complex object from the modulus of its Fourier transform. These results are shown in Appendix E [16]. These results point to further applications of phase retrieval for remote sensing.

2.5 RECONSTRUCTION USING BOUNDARY VALUES

Hayes and Quatieri (H-Q) [35], devised an algorithm for reconstructing a real-valued sampled object from its autocorrelation function assuming that one knew a priori the values of the boundary (edges) of the object. They also claimed that the reconstruction was always unique [35]. In Appendix F [17], four results regarding the H-Q recursive algorithm are described. First, knowledge of the boundary values of the object is not sufficient to ensure a unique reconstruction. Second, even when the solution is ultimately unique, the H-Q recursive algorithm can have ambiguities in intermediate steps. Third, an extension to the H-Q recursive algorithm is given that does find the solution. Fourth, it is demonstrated that the algorithm can be applied to complex-valued objects as well as real-valued objects.

2.6 ESTIMATING FOURIER MODULUS FROM SPECKLE IMAGES

The Labeyrie method (stellar speckle interferometry) of estimating the (squared) modulus of the Fourier transform of the object involves averaging over the squared moduli of the Fourier transforms of many independently degraded speckle images [4]. A new maximum likelihood estimator of the Fourier modulus has been derived. It uses the same data as the Labeyrie method but, using knowledge of the statistics of the atmospheric blurring process, arrives at a better estimate of the Fourier modulus. This estimation procedure is described in Appendix G [18].

2.7 RELAXATION ALGORITHM

Earlier investigations [32] showed that the simplest form of the iterative Fourier transform algorithm, the "error-reduction" algorithm, converged much too slowly to be practical. Improved "input-output" algorithms converge much faster, in fewer than 100 iterations. In seeking still faster convergence, a relaxation-parameter algorithm was devised. It involves solving for the optimum linear combination of two estimates. One version of the relaxation method involves reversing the roles of the two domains -- the object and Fourier domains (then the linear combination is performed in the Fourier domain). This Fourier-domain relaxation approach is better justified mathematically because enforcing the object-domain constraint is a projection in a Hilbert space onto a convex set. A mathematically tractable solution for the optimum linear combination involves the use of a new error metric and the (closed-form) solution of a cubic equation. This is described in some detail in Section 3 of [12]. Unfortunately there was no success making the relaxation algorithm perform even as well as the hybrid input-output version of the iterative Fourier transform algorithm. Part of the reason for the lack of success may be that a linear combination that moves the solution in the direction of the correct Fourier phase typically

yields a result that has too much energy (sum of values squared). One potential solution to this problem would be to redefine the linear combination to be a normalized linear combination. Unfortunately, to find the optimum normalized linear combination one must solve a messy nonlinear equation, which we have not succeeded in doing analytically. Another possibility would be to allow for more than one parameter to define the linear sums. Although the relaxation approach has not yet yielded a useful reconstruction algorithm, the approach still merits further research.

2.8 UNIQUENESS WITH SEPARATED SUPPORT

By using the theory of analytic functions, it was shown that one-dimensional (1-D) objects consisting of sufficiently separated parts are usually uniquely related to their Fourier modulus [19]. This is in sharp contrast to the usual 1-D case for which there are an enormous number of ambiguous solutions. This is described in detail in Appendix B of [12].

2.9 ALTERNATIVE ITERATIVE ALGORITHMS

The theoretical justification for the input-output iterative Fourier transform algorithm [6-8] alludes to a control theory point-of-view. Yet rigorous control theory had not actually been applied to the problem. Two alternative algorithms based on control theory were formulated (see Section 5 of [10]).

In the first, it is assumed that the individual sidelobes of the complex Fourier transform of the object can be modeled by a fairly simple mathematical formula having a small number of free parameters. By curve fitting each lobe of the Fourier modulus to the model, one could determine the parameters and thereby determine the phase. One would first curve fit one larger lobe, compute the magnitude of the model from the fitted parameters, and subtract that model from the

modulus measurement. Smaller lobes would be curvefitted and subtracted resursively from the residual modulus. After all the lobes are modeled, the corresponding phase would be combined with the measured modulus and the image would be computed by inverse Fourier transformation. It is yet to be determined whether the Fourier transform can be modeled as described above.

In the second algorithm, the difference in the phase of the Fourier transform of the current estimate and that of the previous estimate is multiplied by a gain factor and added to the previous phase estimate. This is similar to previous iterative algorithms except that the roles of the two domains are reversed.

Both the methods described above, as well as others, merit further research and implementation.

2.10 SHIFT-AND-ADD

Shift-and-add [36, 37] is a method of reconstructing images of astronomical objects from multiple short-exposure images. It consists of shifting all the images so that their maximum values all lie at the same coordinate, then adding (or averaging) them all. This has been shown to work well for objects consisting of a collection of delta functions (points) [36, 37], but it was not demonstrated for realistic extended objects, such as satellites. We implemented the shift-and-add method on the computer and exercised it on data both for an object consisting of a collection of delta functions and for an extended object.

For the first object, consisting of three delta-functions having relative brightnesses of 10:2:1, the results of applying shift-and-add to 156 severely blurred images resulted in a somewhat blurred, but very recognizable image of the object. We took shift-and-add one step further by combining it with a form of subtractive deconvolution related to the CLEAN method. Then the resulting image was excellent.

A similar experiment was performed for an extended object -- the satellite shown in Figure 7.4-4 of Appendix A. In that case, the reconstructed image resulting from shift-and-add was very poor whether or not CLEAN was used. These results are shown in Section 7 of [10].

This one set of experiments was not sufficient to fully delineate the types of objects for which shift-and-add is effective, but we did demonstrate that shift-and-add works very well for an object consisting of a small number of delta-function-like points dominated by a brightest one, but works very poorly for an extended object, even one containing isolated bright points.

2.11 ANALYSIS OF ASTRONOMICAL DATA

The Harvard-Smithsonian Center for Astrophysics supplied us with two stellar-speckle interferometer data sets, one of a binary star, the other of Jupiter's moon Io. Each data set was recorded with the PAPA photon-counting camera [38] and consists of a stream of detected photon events. Associated with each photon event is a mark denoting the time and location of the event on the detector.

As required for the iterative algorithm, an estimate of the modulus of the Fourier transform of the object is formed by using Labeyrie's method of stellar-speckle interferometry. Each data stream is partitioned in time into a sequence of 10 millisecond exposures; each exposure is then Fourier transformed and averaged. The result is proportional to an estimate of the modulus of the Fourier transform of the object under measurement.

Included in the estimate are the effects of the non-uniform photo response of the camera as well as the modulation transfer function (MTF) of the speckle process. The non-uniform photo response can be corrected to first order by a flat-fielding operation in which each exposure is divided by an image of a uniform flood source. The MTF

is removed by dividing the estimated modulus by an estimated modulus from a point-like reference star, which is an estimate of the modulus of the transfer function of the speckle process.

The data we received included the binary star and Io data plus their respective reference star images and a flat field image. Upon examination of the data, we noticed artifacts in the images which appeared as a burlap-like texture modulated by the actual image content. Figure 1(a) shows the flat field image and Figures 1(b) and 1(c) show long exposure images of the binary star data before and after flat-field correction. Figure 2 shows the same things for the reference star. The sample-standard deviation within a 32 by 32 region of the flat field image was 37 percent; adjacent pixels in this region varied by as much as 250 percent. The sample-standard deviation of the corresponding region of the binary star data was 37 percent before a flat-field correction and 24 percent after a flat-field correction. These artifacts are inherent in the data collection process and discouraged us from investing any further computer time in the reconstruction of images from the Fourier modulus data. Information about a better flat-fielding method was promised by Harvard but was never received. This data deserves further processing, but only after a better flat-fielding is achieved.

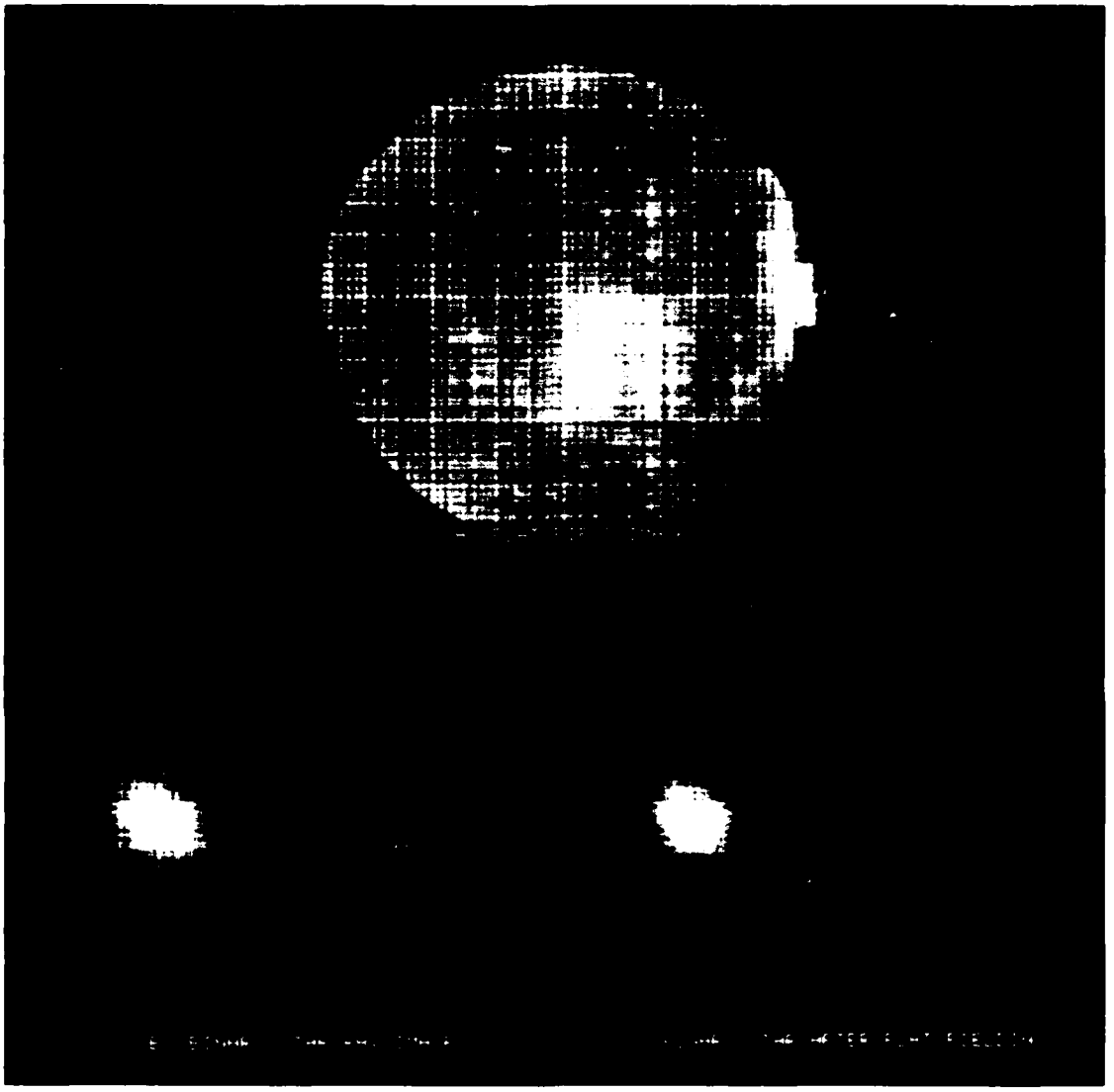


Figure 1. Flat Fielding Experiment with Binary Star Data. (a) Flat field; (b)long-exposure binary star raw image; (c) binary star image after flat field correction.

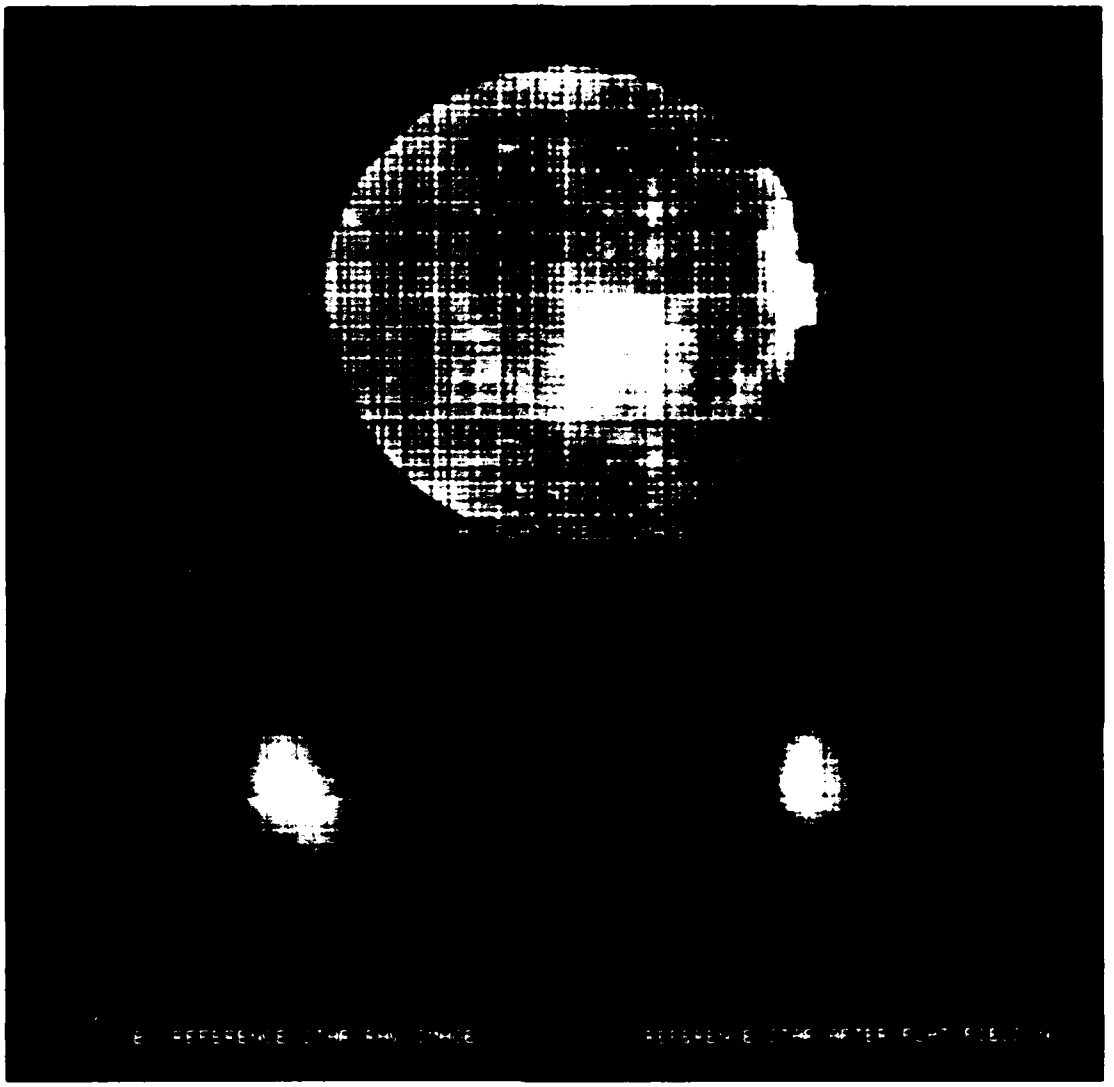


Figure 2. Flat Fielding Experiment with Reference Star Data. (a) Flat field; (b) long-exposure reference star raw image; (c) reference star image after flat field correction.

Appendix A

PHASE RETRIEVAL AND IMAGE RECONSTRUCTION FOR ASTRONOMY

J.C. Dainty

Blackett Laboratory
Imperial College
London SW7 ZBZ ENGLAND

and

J.R. Fienup

Environmental Research Institute of Michigan
P.O. Box 8618, Ann Arbor, MI 48107 USA

Submitted to Image Recovery: Theory and Application,
H. Stark, Ed., Academic Press, Publisher, Chapter 7

CHAPTER 7
PHASE RETRIEVAL AND IMAGE RECONSTRUCTION FOR ASTRONOMY

J.C. Dainty and J.R. Fienup

7.1 Introduction

In optical astronomy, the traditional way of obtaining an object map or object intensity $f(x,y)$[+] is to use a single large telescope to record a conventional long-exposure image. This image can be enhanced, if necessary, using the standard techniques of digital image processing. Although this may, at first sight, seem to be the most obvious method of estimating the object intensity, it suffers from two drawbacks. First, for Earth-based observation, atmospheric turbulence limits the resolution obtainable to approximately one arc-second (at the best observing sites) compared to the limit set by diffraction, $\alpha_{min} \sim \lambda/D$, where λ is the wavelength and D is the telescope diameter. This limit is approximately 0.02 arc-sec for a 5m telescope at a wavelength of 400 nm. Second, ignoring atmospheric turbulence, the diameter D of a single large telescope limits the resolution (as above) and it is difficult to imagine a telescope with a diameter greater than about 50m in the forseeable future.

The solution to the problem of limited resolution lies with a family of techniques which can be described as interferometric imaging. The basis of these is the van Cittert-Zernike theorem [1, Sections 5.6 and 7.4], which, in the form of use in astronomy, states that the spatial coherence function $F(u,v)$ in the far-field of a thermal light source is proportional to the Fourier transform of the object intensity $f(x,y)$. The spatial coherence of starlight can be measured using a variety of techniques, including speckle and Michelson interferometry.

Some of these interferometric methods are particularly suited to long baseline and synthetic aperture techniques, in which the

[+]In astronomy, it is customary to use angular variables (α,β) instead of the spatial variables (x,y) used in this chapter.

"diameter" D is potentially 1 km or more, providing the possibility of sub-milliarcsecond resolution in the future. Synthetic aperture techniques are of course widely used for imaging in radio-astronomy.

Interferometric imaging techniques in optical astronomy all suffer from one major disadvantage: it is difficult, and sometimes impossible, to get an accurate estimate of the phase of $F(u,v)$. We are then faced with the problem of reconstructing the object intensity $f(x,y)$ either from $|F(u,v)|$ alone or with limited phase information. In both cases, the modulus (and phase information, if any) is known only within a finite region of the spatial frequency (u,v) plane, typically in a band extending from 0 to $\sqrt{u_{max}^2 + v_{max}^2} = D/\lambda$. In addition, $f(x,y)$ is known to be a nonnegative function and of finite extent.

In Section 7.2 we summarize the uniqueness aspects of phase retrieval from modulus-only data (this is discussed in Chapter 6 in more detail). Section 7.3 describes several algorithms for modulus-only data, both for special and general objects. Iterative algorithms have proved to be the most successful in practice and Section 7.4 describes these and gives detailed instructions on how to obtain satisfactory reconstructions using a particular iterative approach. In the final section, we describe a number of solutions to the phase problem in optical astronomy that are specific to speckle and pupil plane (Michelson) interferometry. All of these utilize the (limited) phase information provided by these particular observing techniques.

Other reviews of stellar speckle interferometry can be found in References 2-4.

7.2 Uniqueness of Phase Retrieval from Modulus Data

In interferometric imaging, an attempt is made to measure the object's Fourier transform

$$F(u,v) = |F(u,v)| \exp [\exp i\psi(u,v)]$$

$$= \int\int_{-\infty}^{\infty} f(x,y) \exp [-i2\pi(ux + vy)] \, dx \, dy, \qquad (7.2\text{-}1)$$

where $f(x,y)$ is the object's brightness or intensity distribution, $|F(u,v)|$ is the Fourier modulus and $\psi(u,v)$ is the Fourier phase. In the extreme case of severe atmospheric turbulence, which is ordinarily the case at optical wavelengths, only $|F(u,v)|$ can be measured faithfully and all information about the phase, $\psi(u,v)$, is lost. Then two questions arise: first, how can one reconstruct $f(x,y)$, or equivalently reconstruct $\psi(u,v)$, from $|F(u,v)|$ given the a priori information that a physical brightness distribution is a real, non-negative function? This first question is answered in Sections 7.3 and 7.4 which discuss various reconstruction algorithms. The second question is, given an image produced by a reconstruction algorithm, how does one know whether it is unique? That is, might not there be other real, non-negative solutions having the same Fourier modulus?

First it is necessary to recognize that $f(x,y)$, $f(x - x_0, y - y_0)$, and the image rotated by 180° (reflected through the origin), $f(-x - x_0, -y - y_0)$, all have the same Fourier modulus. This fundamental ambiguity does not cause much concern since it is the appearance of the object that is usually sought rather than its position. The position and the 180° rotational ambiguity can be easily resolved in many cases from a very low-resolution image of the object. If the above fundamental ambiguities are the only ambiguities, we refer to the object as being unique; only ambiguous solutions having different forms than the object are considered ambiguous.

Until less than one decade ago, there was much pessimism that the phase retrieval problem could be solved or that the solution would be useful, because the available theory at the time, which was for one-dimensional functions, indicated that there were hopelessly many ambiguous solutions [5]. It can be shown that Fourier transforms of one-dimensional functions of finite support are analytic functions of exponential type. They are characterized by their complex zeros, i.e., the locations of the zeros of the (Fourier) function analytically extended into the complex plane, by

$$F(z) = \int_{-\infty}^{\infty} f(x)\, e^{-i2\pi zx} dx, \qquad (7.2\text{-}2)$$

where $z = u + iu'$, u being the physical real axis and u' being the imaginary axis. If the set of complex zeros, for which $F(z) = 0$, are $\{z_n\}$, then the Fourier transform can be expressed as the Hadamard product

$$F(z) \propto \prod_{n=1}^{\infty} (1 - z/z_n). \qquad (7.2\text{-}3)$$

Given $|F(u)|^2 = F(u)F^*(u)$, the complex zeros of its analytic extension into the complex plane, $|F(z)|^2 = F(z)F^*(z^*)$, can be computed, but they include both the zeros, $\{z_n\}$, of $F(z)$ and the zeros of $F^*(z^*)$, which are $\{z_n^*\}$.

If $F(z)$ has N complex zeros, then not knowing which of the 2N complex zeros of $|F(z)|^2$ go with $F(z)$ and which go with $F^*(z^*)$ leads to a 2^N-fold ambiguity. (Half of this number of solutions are the mirror images of the other half, reducing the ambiguity to 2^{N-1}-fold.) Ordinarily, N is comparable to the space-bandwidth product of $f(x)$, and so 2^N is an enormous ambiguity, making the one-dimensional problem seem hopeless. Only for the case of objects known to consist of two separated parts is the one-dimensional phase retrieval problem likely to be unique and therefore interesting [6-8].

The first indications that the situation is much different in two dimensions came from empirical reconstruction experiments [9, 10]: images that were reconstructed looked like the simulated objects used to compute the Fourier modulus data. Those results gave hope that the two-dimensional problem may be soluble and unique, giving impetus to attempts to extend the one-dimensional theory of analytic functons to two dimensions in such a way that predictions concerning uniqueness could be made . Unfortunately, that theoretical level of understanding of the two-dimensional problem, while progressing [11-16] has not matched that available for the one-dimensional case.

If the object is approximated by an array of delta functions or by an array of numbers in a computer, then its Fourier transform reduces to a discrete Fourier transform which, with a change in variables, can be written as a polynomial of a complex variable. When this discrete approximation is made, then one can understand the vast difference between the one-dimensional and two-dimensional phase retrieval problems [16]. It can be shown in the one-dimensional case that the zeros of the polynomial are analogous to the complex zeros for the continuous (analytic functions) case, and that there are 2^{N-1} ambiguous solutions where N is the number of factors into which the polynomial can be factored. Since the fundamental theorem of algebra states that a polynomial of degree N has N prime (irreducible) factors, one has a high degree of ambiguity for the one-dimensional case. As was pointed out by Bruck and Sodin [16], it is known that polynomials of two complex variables are rarely factorable. Hence the two-dimensional phase retrieval problem is usually unique (ignoring noise issues). Chapter 6 describes the uniqueness issues for discrete functions in some detail.

Although most two-dimensional discrete objects are unique, one would like to know what conditions ensure uniqueness. As described in Section 6.3, objects whose Fourier transforms satisfy Eisenstein's

criterion are unconditionally unique [17]. The type of support that the object must have (i.e., the set of points or the shape within which the object is nonzero) to satisfy this criterion is the only known support which guarantees unconditional uniqueness. The uniqueness of such objects can also be demonstrated by the existence of a recursive algorithm (a closed-form solution) that can reconstruct such objects from their Fourier modulus via their autocorrelation functions [18]. This recursive algorithm can also be used to reconstruct in closed form objects having other supports as well, such as triangular supports having nonzero corners. But for these other objects the support must be known a priori, since there may be other solutions having different supports [18]. Objects that are unique among the class of objects having that given support include all objects for which the convex hull of their support is a polygon having no parallel sides [19, 20]. Another such class of objects are those satisfying the holography condition: the object consists of an extended part plus a delta function (unresolved component) separated from the extended part by a distance greater than the diameter of the extended part. Then the extended part of the object is found as an isolated term in the autocorrelation function. Note, however, that objects satisfying the holography condition are not necessarily unique among all objects -- they are only guaranteed to be unique among the class of objects satisfying the holography condition.

Although it remains to be proven definitively, the general feeling is that the physical case of interest, the two dimensional continuous case, mirrors the discrete case and will, except for special objects, usually (but not always) be unique. It does not appear that the addition of small amounts of noise will change that assessment [21].

7.3 Algorithms for Phase Retrieval from Modulus

Since most phase retrieval/image reconstruction algorithms are performed on sampled data in the computer, we will employ the discrete Fourier transform

$$F(u,v) = |F(u,v)| \exp [i\psi(u,v)] = \mathcal{F}[f(x,y)]$$

$$= \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \exp [-i2\pi(ux + vy)/N] \qquad (7.3\text{-}1)$$

and its inverse

$$f(x,y) = N^{-2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v) \exp [i2\pi(ux + vy)/N] \qquad (7.3\text{-}2)$$

where u, v, x and y = 0, 1, 2, . . ., N - 1 (for simplicity square arrays are assumed). Any computation of the discrete Fourier transform is of course accomplished using the Fast Fourier Transform (FFT) method. Note that in order to compute $|F(u,v)|^2$ without aliasing it is required that f(x,y) be zero outside of $0 \leq x \leq M - 1$ and $0 \leq y \leq M - 1$, where $M \leq N/2$. The autocorrelation of f(x,y) is given by

$$r_f(x,y) = \sum_{x'=0}^{N-1} \sum_{y'=0}^{N-1} f(x',y')f(x' - x, y' - y) \qquad (7.3\text{-}3)$$

$$= \mathcal{F}^{-1}[|F(u,v)|^2]. \qquad (7.3\text{-}4)$$

The problem of reconstructing $\psi(u,v)$ or f(x,y) from |F(u,v)| or from $r_f(x,y)$ has received much attention, especially within the last decade. Section 7.3A describes several reconstruction methods that are tailored to special types of objects. Most of these methods work only for objects consisting of a collection of delta-function-like points or at least containing a point-like component. Although

they only work for such special cases, they are of interest because in astronomy the existence of point-like stars is common and these methods tend to be simpler than the methods for more general objects. Section 7.3B describes more general methods designed to work on most types of astronomical objects (objects of finite extent). Section 7.4 describes in more detail the iterative Fourier transform algorithm, which as been the most general and most successful reconstruction algorithm to date.

A Algorithms for Special Types of Objects

(i) Holography

The classic reconstruction method, discussed briefly in Section 7.2, is holography [22]: when the object includes a delta-function component (an unresolved point called the reference point) sufficiently separated from the rest of the object, then the object can be found as one term in the autocorrelation. Although holography was originally conceived for applications employing coherent light, the same principles apply for incoherent light if one has measured the modulus of the Fourier transform of the object [23, 24].

(ii) Recursive Reconstruction Using Latent Reference Points

The closed-form recursive method [18] mentioned in Section 7.2 is akin to holography. It requires the existence of two or more reference points, making it more restrictive than the holography condition, but these reference points can be very close to the extended part of the object, making it less restrictive than the holography condition. These reference points satisfy the holography condition only for the far edge of the object which is immediately reconstructed from the autocorrelation. After those edge points are solved (including the reference points), then at least one of the autocorrelation values can be expressed as a sum of products of the previously solved points plus the product of an unknown point with a

reference point. This single linear equation in a single unknown may be solved trivally. The rest of the points in the object are solved recursively in a similar manner. The reference points are refered to as latent since they cannot be used to solve for the interior points in the object until the exterior points are first reconstructed. This algorithm works only for particular types of sampled objects, including those for which the convex hull of their support is a polygon having no parallel sides[+] [20]. As mentioned in Section 7.2, one must know the support a priori to employ this algorithm. Except in special cases, given just the support of the autocorrelation one cannot find the support of the object [25]: one can only narrow down the possibilities.

(iii) Difference Fourier Synthesis

Some of the phase retrieval methods from x-ray crystallography [26] can be used for the astronomical application. One of these is difference Fourier synthesis [27]. Suppose the object consists of a collection of point-like stars and one has an estimate of the object that contains some, but not all, of the object's stars. One creates a new image by (1) Fourier transforming the estimate of the object, (2) replacing the computed Fourier modulus by the true Fourier modulus while keeping the computed Fourier phase, and (3) inverse Fourier transforming the result. Under the assumptions mentioned above, it can be shown that the new image will contain the stars missing in the object's estimate, but at half their true value. So by taking the difference between the new image and the estimate of the object one can identify the locations and relative brightnesses

[+]Note that Figure 6(a) in [18] has two solutions rather than the single solution claimed. It would have a single solution if either point C' or point C" were zero.

of the missing stars and use that information to form a new estimate of the object. This algorithm must be applied iteratively because extraneous terms are produced in the new image along with the desired missing stars. This method was successful for point-like objects but not for extended objects [27]. It is tantalizingly close to the iterative transform algorithm described in Section 7.4 which is applicable to a much more general class of objects.

(iv) Products of Autocorrelations

If the object consists of a collection of point-like stars, then the autocorrelation, $r_f(x,y)$, also consists of a number of point-like terms. First one computes the product $r_f(x - x_0, y - y_0) \cdot r_f(x,y)$, where the point (x_0,y_0) is one for which $r_f(x_0,y_0)$ is nonzero. Next one computes the triple product $r_f(x - x_1, y - y_1)\, r_f(x - x_0, y - y_0)\, r_f(x,y)$, where the point (x_1,y_1) is one for which $r_f(x_1 - x_0, y_1 - y_0) \cdot r_f(x_1,y_1)$ is nonzero. If the object points satisfy certain nonredundancy conditions, then the support of the triple product is identical to the support of the object (or the object rotated by 180°). Furthermore, the values of the object can be reconstructed in closed form by some very simple equations involving taking the cube root of (all but three of) the points in the triple product [25]. The nonredundancy conditions that the object must satisfy are that (1) the vector separations between three pairs of six distinct points must not sum to zero, (2) the vector separations between two different pairs of points must not be equal, and (3) the vector separation between one pair of points must not equal twice the vector separation between another pair of points. It is likely that a randomly-distributed collection of stars would satisfy the nonredundancy conditions unless the number of stars in the image is large.

B Methods Proposed for General Objects

The phase retrieval problem for general objects of finite support can be approached in a number of ways. A brute-force searching through all possible objects or all possible phases is, of course, impossible. For example if there are K different values for each of the M^2 pixels of f(x,y) that must be searched to see which objects agree with the given Fourier modulus, then the number of possibilities to be investigated is K^{M^2}, an overwhelming number, even for K = 2. If each of the particles in the universe were a computer that could search 10^9 objects per second, then it would take the age of the universe to search all possible binary objects for M = 19. Clearly a very intelligent algorithm is needed to search through the possible objects to find one consistent with the Fourier modulus data.

Consider taking the modulus squared of both sides of Eq. (7.3-1) or simply use Eq. (7.3-3). In either case one has a system of N^2 nonlinear equations -- one for each value of $|F(u,v)|^2$ or of $r_f(x,y)$ — in M^2 unknowns, the values of f(x,y) within its support. One could equally well treat the Fourier phases as the unknowns. Obviously there is a high degree of redundancy in the set of equations since there are more equations than unknowns. Nevertheless for large values of M and N the solution of a set of such nonlinear equations is a formidable task.

The classic approach to such a problem is to define an error metric for any given estimate for the object (or for the phase) and use a Newton-Raphson or a gradient search method to seek a solution, that is, an estimate for which the error metric is zero. For example, the error metric could be

$$E_r = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} [r_g(x,y) - r_f(x,y)]^2 \qquad (7.3\text{-}5)$$

where $r_g(x,y)$ is the autocorrelation of $g(x,y)$, an estimate of $f(x,y)$. If $E_r = 0$ for a $g(x,y)$ that satisfies the object-domain constraints, then $g(x,y)$ is a solution. The solution is necessarily equal to $f(x,y)$ only if the phase retrieval problem is unique for $f(x,y)$, but at the reconstruction stage of the game this is not a problem since we are only concerned with finding a solution (or solutions).

The error metric E_r can be thought of as a function in an M^2-dimensional space where the values $g(x,y)$ are the coordinates of the space. The collection of the M^2 values of $g(x,y)$ can be thought of as an M^2-dimensional vector coordinate in that space. The problem is to find, within the subset of that space for which $g(x,y)$ satisfies the object-domain nonnegativity and support constraints, the point at which E_r is at a global minimum ($E_r = 0$ if no noise is present). One can seek the global minimum by starting with an estimate, $g_0(x,y)$, computing the gradient of E_r at that point, and use that information to step down the side of the hill toward a lower value of E_r at a new estimate $g_1(x,y)$. Further steps are taken iteratively until the global solution is found or until stagnation occurs at a local minimum.

(i) Newton-Raphson

The Newton-Raphson method [28] involves the local linearization of the problem by expressing $r_g(x,y)$ in Eq. (7.3-5) in a Taylor series expansion about the current estimate, $g_k(x,y)$:

$$g(x,y) = g_k(x,y) + \Delta g(x,y), \tag{7.3-6}$$

and keeping only up to first-order terms:

$$r_g(x,y) = r_{g_k}(x,y) + \sum_{x'} \sum_{y'} \frac{\partial r_g(x,y)}{\partial g(x',y')} \Delta g(x',y'). \tag{7.3-7}$$

Inserting this into the expression for E_r, E_r is minimized by setting equal to zero the partial derivatives of E_r with respect to $\Delta g(x_o,y_o)$ for each value of (x_o,y_o). This gives a system of M^2 linear equations in M^2 unknowns (the Δg's) which can be solved by the inversion of an M^2 by M^2 matrix. A new estimate is given by Eq. (7.3-6) and the procedure is repeated. The matrix inversion takes on the order of M^6 operations, making this method computationally very intensive despite the fact that only a small number of iterations are required.

(ii) Gradient Search

Although requiring several times as many iterations as the Newton-Raphson method, gradient search algorithms are attractive because the amount of computation per iteration is far less. Ordinarily the computation of the gradient itself, if done using finite differences, would be very time consuming, requiring one or two times M^2 FFT's. However, by a proper choice of the error metric it is possible to compute the entire gradient using only two FFT's [29]. If the error metric is defined in the Fourier domain by

$$E_F^2 = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} [|G(u,v)| - |F(u,v)|]^2, \qquad (7.3\text{-}8)$$

where $G(u,v)$ is the Fourier transform of the current estimate $g(x,y)$, then it can be shown that the bulk of the computation for the gradient can be accomplished by the first three steps of the iterative transform algorithm, which is described in Section 7.4.

In general, the kth iteration of a gradient search method would proceed by computing a desired direction, $d_k(x,y)$, in which to go from a given object estimate, $g_k(x,y)$, and take a step of size h_k in that direction:

$$g_k''(x,y) = g_k(x,y) + h_k\, d_k(x,y) \qquad (7.3\text{-}9)$$

For the steepest descent gradient search method, $d_k(x,y)$ is just the negative of the partial derivative of E_F^2 with respect to $g(x,y)$. For the conjugate gradient method, $d_k(x,y)$ is a linear sum of the negative of the gradient and $d_{k-1}(x,y)$. After stepping to $g_k''(x,y)$ a test is made to see if it violates the object-domain constraints. It usually does, and a new estimate $g_{k+1}(x,y)$ is taken to be the function in object space satisfying the object-domain constraints that is closest to $g_k''(x,y)$. [In the language of Chapter 8, $g_k''(x,y)$ is projected onto the space of functions satisfying the constraints.] As for most applications, the steepest descent method is very slow to converge, and the conjugate gradient method is much better [29]. Although the gradient search methods have not proved to converge as quickly as the algorithm described in Section 7.4, they have not yet been fully optimized and deserve further study.

(iii) Tracking of 1-D Complex Zeros

The method of tracking 1-D complex zeros [10] starts with one-dimensional radial cuts through the Fourier modulus. By the projection-slice theorem a radial cut through the 2-D Fourier transform is the 1-D Fourier transform of the projection of the object along a direction perpendicular to the direction of the radial cut. As mentioned in Section 7.2, there are 2^N different solutions for a given 1-D radial cut, all of which can be computed by using different combinations of the conjugated or nonconjugated zeros $\{z_n, z_n^*\}$. Of these solutions only a small fraction will correspond to nonnegative projected images, and only those need be retained since the object brightness distribution is known to be nonnegative. As the angle of the cut is rotated slightly, the complex zeros should change in a continuous fashion, and so the complex zeros for one cut can be associated with the complex zeros for other cuts. By finding the complex zeros that yield nonnegative projected images over a range of cut angles, one can identify the set

of zeros $\{z_n\}$ associated with each cut of the Fourier transform and solve the 1-D phase retrieval problem for each of the cuts, thereby solve the 2-D phase retrieval problem. This method has the disadvantage of having to compute 2^N different images for a number of cuts, which is computationally very burdensome. Application of this method gave one of the first indications that the 2-D problem is more likely to be unique than the 1-D problem [10].

(iv) Making 1-D Complex Zeros Consistent Between Orthogonal Cuts

Employing a similar philosophy to the method described above, this method [30] computes the complex zeros of 1-D cuts through the 2-D Fourier modulus squared, and then from the complex zeros computes all 2^N different phase solutions for each cut. In this case the cuts are each line and each column of the 2-D Fourier transform. At each point in the Fourier domain a true solution must have a phase that is consistent with the phase of both orthogonal cuts. By checking consistency in phase at each point amongst all 2^N possible phases for each cut, it is possible to find what 2-D phase functions are consistent across the entire Fourier domain. A disadvantage of this method is the enormous computational burden of computing 2^N solutions for each of 2N cuts and comparing all the potential combinations.

(v) Sampling Theorem Methods

The use of the sampling theorem is based on knowing that the object is of finite extent. If the object is zero outside a square of length L centered at the origin, then the Shannon sampling theorem gives us

$$F(u,v) = \sum_p \sum_q F(p\Delta u, q\Delta v)\ \mathrm{sinc}\left(\frac{u - p\Delta u}{\Delta u}\right) \mathrm{sinc}\left(\frac{v - q\Delta v}{\Delta v}\right) \quad (7.3\text{-}10)$$

where $\Delta u = \Delta v = L^{-1}$. In the sampling theorem method [31, 32], the samples $F(p\Delta u, q\Delta v)$ are estimated by $|F(p\Delta u, q\Delta v)| \exp [i\phi(p\Delta u, q\Delta v)]$, where ϕ is an estimate of the unknown phase. This is inserted into the right-hand side of Eq. (7.3-10), which is evaluated at the half-integer coordinates, and the squared modulus is computed. Since adequate sampling of $|F(u,v)|^2$ is twice the sampling required for $F(u,v)$, and it is assumed that $|F(u,v)|^2$ was adequately sampled, the computed squared modulus values can be compared with the measured value of $|F(u,v)|^2$. This leads to a system of N^2 nonlinear equations in N^2 unknowns which can be solved, for example, by the Newton-Raphson algorithm. This differs from the Newton-Raphson method discribed in Section 7.3B(i) -- the set of nonlinear equations to be solved are different -- but the computational complexity is the same.

A computationally more efficient version of the sampling theorem method has been devised [33]. Letting $v = 0$, Eq. (7.3-10) reduces to

$$F(u,0) = \sum_p F(p\Delta u, 0) \operatorname{sinc} \left(\frac{u - p\Delta u}{\Delta u}\right). \qquad (7.3\text{-}11)$$

Since the half-integer values of u along the $v = 0$ axis depend only on the sampled values along the $v = 0$ axis, the phase of $F(u,0)$ can be determined by solving the one-dimensional problem which involves only N nonlinear equations in N unknowns. In a similar fashion, given $F(p\Delta u, 0)$ for each integer value of p the phase of $F(p\Delta u, v)$ can be determined by solving a system of N nonlinear equations in N unknowns. The total required computation to solve $(N + 1)$ sets of N nonlinear equations in N unknowns, when solved by the Newton-Raphson method, involves $(N + 1)N^3$ operations times the number of iterations per 1-D solution, which is a great savings compared with the fully 2-D form described earlier. A disadvantage of this method is that each of the 1-D problems that is solved is just one of the 2^N ambiguous solutions that are always present for the 1-D

problem. Consequently this method will fail unless one starts with an initial estimate very close to the true solution or until one finds a way of incorporating more constraints into the algorithm.

A third variation on the sampling theorem method is the crude phase estimation method [34]. If one approximates the sinc (u) kernels in Eq. (7.3-10) by a function that is unity at $u = 0$, a constant β at $u = \pm 1/2$, and is zero for all other integer and half-integer values of u, then Eq. (7.3-10) simplifies to such an extent that the phase can be expressed in closed form as a function of the modulus of Eq. (7.3-10). The closed-form expression involves an arc-cosine function which leads to two-fold sign ambiguity, which can be resolved by a consistency check over a closed 2-D path in Fourier space [34]. While this method is computationally simple, it unfortunately yields only a very crude estimate of the Fourier phase owing to the crudeness with which the sinc function is approximated.

7.4 Iterative Transform Algorithm

The iterative transform algorithm is a descendant of the Gerchberg-Saxton algorithm [35-37]. It bounces back and forth between the object domain, where a priori knowledge about the object -- its nonnegativity and bounds on its support -- are applied, and the Fourier domain, where the measured Fourier modulus data is applied. The iterative transform algorithm differs from the Gerchberg-Saxton algorithm both in the type of data and constraints available and in the more powerful object-domain operations applied for the hybrid input-output version of the algorithm. The iterative transform algorithm has been the most successful method for reconstructing nonnegative objects from the modulus of their Fourier transforms because it works for the most general types of objects, uses all the available data and constraints, is not highly sensitive to noise, and is not as computationally burdensome as most other methods. In Section 7.4A the basic iterative algorithm is

described. Additional details that will be helpful to the practical implementation of the algorithm are included in Section 7.4B. In some difficult cases the basic iterative transform algorithm fails to converge to a solution. Distinct modes of stagnation have been identified and methods for overcoming them have been devised, as described in Section 7.4C. Section 7.4D shows an example of a simulation of noisy stellar speckle inteferometry data, Fourier modulus estimation and iterative transform reconstruction.

A The Basic Iterative Algorithm

(i) The Error-Reduction Algorithm

The simplest version of the iterative transform algorithm follows the philosophy of the Gerchberg-Saxton algorithm [35-37], and it is known as the error-reduction algorithm [9, 29, 38]. It can be viewed in a number of different ways: in terms of the method of successive approximations [39], as a form of steepest-descent gradient search [29], or as a projection onto sets in a Hilbert space (the Fourier modulus constraint being onto a nonconvex set, however, so convergence is not assured). The latter point of view is described in more detail in Chapter 8.

For the most general problem, the error-reduction algorithm consists of the following four steps (for the k^{th} iteration): (1) Fourier transform $g_k(x,y)$, an estimate of $f(x,y)$, yielding $G_k(u,v)$; (2) make the minimum changes in $G_k(u,v)$ which allow it to satisfy the Fourier-domain constraints to form $G'_k(u,v)$, an estimate of $F(u,v)$; (3) inverse Fourier transform $G'_k(u,v)$, yielding $g'_k(x,y)$, the corresponding image; and (4) make the minimum changes in $g'_k(x,y)$ that allow it to satisfy the object-domain constraints to form $g_{k+1}(x,y)$, a new estimate of the object. For phase retrieval for a nonnegative object from the Fourier modulus $|F(u,v)|$, these four steps are

Step 1: $G_k(u,v) = |G_k(u,v)| \exp[i\phi_k(u,v)] = \mathcal{F}[g_k(x,y)]$, (7.4-1)

Step 2: $G'_k(u,v) = |F(u,v)| \exp[i\phi_k(u,v)]$, (7.4-2)

Step 3: $g'_k(x,y) = \mathcal{F}^{-1}[G'_k(u,v)]$, (7.4-3)

Step 4: $$g_{k+1}(x,y) = \begin{cases} g'_k(x,y), & (x,y) \notin \gamma \\ 0, & (x,y) \in \gamma \end{cases}, \quad (7.4\text{-}4)$$

where γ is the set of points at which $g'_k(x,y)$ violates the object-domain constraints and where g_k, G'_k and ϕ_k are estimates of f, F and the phase ψ of F, respectively. The algorithm is typically started by using an array of random numbers for $g_0(x,y)$ or for $\phi_0(u,v)$. Figure 7.4-1 shows a block diagram of the iterative transform algorithm.

For the astronomy problem, the object-domain constraints are the object's nonnegativity and a (usually loose) support constraint. Then the points in the set γ are those outside the assumed support and those within the support for which $g'_k(x,y) < 0$. The diameter of the object can be computed since it is just half the diameter of the autocorrelation; however, the exact support of the object in general cannot be determined uniquely from the support of the autocorrelation [25], and so the support constraint cannot be applied tightly. For other problems, one may not have a nonnegativity constraint, but have a priori knowledge of a tighter support constraint [40].

For the problem of phase retrieval from two intensity measurements [i.e. $|F(u,v)|^2$ and $|f(x,y)|^2$], $g'_k(x,y) = |g'_k(x,y)| \exp[i\theta'_k(x,y)]$ is complex-valued, and Step 4 becomes

$$g_{k+1}(x,y) = |f(x,y)| \exp[i\theta_{k+1}(x,y)] = |f(x,y)| \exp[i\theta'_k(x,y)], \quad (7.4\text{-}5)$$

where $|f(x,y)|$ is the known modulus of the complex-valued object and θ_k is an estimate of the phase of the object. With this modulus

constraint in the object domain, the error-reduction algorithm is precisely the Gerchberg-Saxton algorithm. In this chapter we consider only the problem of phase retrieval from a single intensity measurement, the squared Fourier modulus.

A measure of the convergence of the algorithm to a solution (a Fourier transform pair satisfying all the constraints in both domains) is the normalized root-mean-squared-error (NRMS) metric in the Fourier domain,

$$E_F = \left[\frac{\sum_u \sum_v [|G(u,v)| - |F(u,v)|]^2}{\sum_u \sum_v |F(u,v)|^2} \right]^{1/2}, \qquad (7.4\text{-}6)$$

or in the object domain,

$$E_o = \left[\frac{\sum\sum_{(x,y)\epsilon\gamma} |g_k'(x,y)|^2}{\sum_x \sum_y |g_k'(x,y)|^2} \right]^{1/2} \qquad (7.4\text{-}7)$$

where γ is defined as in Eq. (7.4-4). Unless otherwise noted, the summations are performed over all points in image or Fourier space.

(ii) Proof of Convergence for the Error-Reduction Algorithm

It can be shown that the error-reduction algorithm "converges" in the weak sense that the squared error cannot increase with an increasing number of iterations [29]. This can be shown by considering an unnormalized squared error at the kth iterations,

$$e_{Fk}^2 = N^{-2} \sum_u \sum_v [|G_k(u,v)| - |F(u,v)|]^2 \qquad (7.4\text{-}8)$$

$$= N^{-2} \sum_u \sum_v |G_k(u,v) - G_k'(u,v)|^2 \qquad (7.4\text{-}9)$$

which by Parseval's theorem is

$$e_{Fk}^2 = \sum_x \sum_y |g_k(x,y) - g_k'(x,y)|^2. \qquad (7.4\text{-}10)$$

Now compare this with the corresponding expression for the object-domain error

$$e_{ok}^2 = \sum\sum_{(x,y)\epsilon\gamma} [g_k'(x,y)]^2 \qquad (7.4\text{-}11)$$

$$= \sum_x \sum_y |g_{k+1}(x,y) - g_k'(x,y)|^2 \qquad (7.4\text{-}12)$$

Both $g_k(x,y)$ and $g_{k+1}(x,y)$ by definition satisfy the object-domain constraints. Also at any point (x,y), by definition $g_{k+1}(x,y)$ is the nearest value to $g_k'(x,y)$ that satisfies the object-domain constraints. Therefore for all points,

$$|g_{k+1}(x,y) - g_k'(x,y)| \leq |g_k(x,y) - g_k'(x,y)| \qquad (7.4\text{-}13)$$

and therefore, from Eqs. (7.4-10) and (7.4-12),

$$e_{ok}^2 \leq e_{Fk}^2. \qquad (7.4\text{-}14)$$

Similarly, by Parseval's theorem

$$e_{ok}^2 = N^{-2} \sum_u \sum_v |G_{k+1}(u,v) - G_k'(u,v)|^2 \quad . \qquad (7.4\text{-}15)$$

Since both $G_k'(u,v)$ and $G_{k+1}'(u,v)$ satisfy the Fourier-domain constraint, having modulus $|F(u,v)|$, and by definition $G_{k+1}(u,v)$ is the nearest value to $G_{k+1}(u,v)$ that satisfies the Fourier-domain constraint, then

$$|G_{k+1}(u,v) - G_{k+1}'(u,v)| \leq |G_{k+1}(u,v) - G_k'(u,v)| . \qquad (7.4\text{-}16)$$

Therefore from Eqs. (7.4-9) and (7.4-15),

$$e^2_{F,k+1} \leq e^2_{ok}, \tag{7.4-17}$$

and combining this with Eq. (7.4-14) gives the desired result

$$e^2_{F,k+1} \leq e^2_{ok} \leq e^2_{Fk}. \tag{7.4-18}$$

That is, the error can only decrease, or at worst stay the same, at each successive iteration of the error-reduction algorithm.

(iii) The Hybrid Input-Output Algorithm

Although it works well for the problem of phase retrieval from two intensity measurements, the error-reduction algorithm usually converges very slowly for the problem of phase retrieval from a single intensity measurement being considered here [29]. Several modifications of the iterative transform algorithm were made and tested, and most of them converged faster than the error-reduction algorithm [29]. To date, the most successful version is the hybrid input-output algorithm, which replaces Step 4 of the algorithm by [9, 29]

$$g_{k+1}(x,y) = \begin{cases} g'_k(x,y) & (x,y) \notin \gamma \\ g_k(x,y) - \beta g'_k(x,y), & (x,y) \in \gamma, \end{cases} \tag{7.4-19}$$

where β is a constant feedback parameter. Values of β between 0.5 and 1.0 work well. When using the hybrid input-output algorithm, $g_k(x,y)$ is no longer an estimate of $f(x,y)$; it is instead the input function used to drive the output $g'_k(x,y)$ [which is an estimate of $f(x,y)$] to satisfy the constraints. Hence only the object-domain error E_o is meaningful. When using the hybrid input-output algorithm, even E_o does not always correlate with image quality as well as one would like. For this reason one may prefer to perform a number of cycles of iterations, where one cycle consists of, say, 20 to 50 iterations of the hybrid input-output algorithm followed by 5 to 10 iterations of the error-reduction algorithm, and note E_o only at the end of a cycle.

For a more complete description of the various versions of the iterative algorithm, see Reference 29. Additional details concerning the implementation of the algorithm are given in Section 6. A description of the algorithm as it applies to a number of different problems is given in Reference 39.

B Additional Aspects of the Iterative Transform Algorithm

In the past, some researchers have had varying success in applying the iterative transform algorithm to phase retrieval from a single intensity measurement. In this section, a number of additional aspects of making the iterative algorithm work are given as an aid to the practical implementation of the algorithm.

The data one must have available is an estimate, $|\hat{F}(u,v)|$, of the modulus, $|F(u,v)|$, of the Fourier transform of the object. Although the iterative transform reconstruction algorithm is not hypersensitive to noise, care must be taken to obtain the best possible estimate of the Fourier modulus, which may involve considerable compensation of the raw data [41], depending on how it is collected. In many circumstances one can estimate the expected value of the normalized root-mean-squared (NRMS) error of the data:

$$E_{|\hat{F}|} = \left[\frac{\sum_u \sum_v [|\hat{F}(u,v)| - |F(u,v)|]^2}{\sum_u \sum_v |F(u,v)|^2} \right]^{1/2} . \quad (7.4\text{-}20)$$

As described later, this is useful for deciding when one is close enough to a solution.

As the iterations progress, the NRMS error in the object domain, E_o, given by Eq. (7.4-7), should be computed. E_o is a measure

of how close the current Fourier transform pair is to a solution. Note that the denominator of Eq. (7.4-7) is a constant that need be computed only once. When E_0 goes significantly below $E_{|\hat{F}|}$ one has a solution consistent with the measured data and constraints to within the limits of the error in the given data. It is unlikely that E_0 will ever go to zero because (1) diffraction by the telescope aperture causes sidelobes in the image that extend beyond the support of the object and (2) noise in $|\hat{F}(u,v)|$ almost always results in a Fourier modulus that is inconsistent with either the nonnegativity constraint or any reasonable support constraint. Furthermore, in the presence of noise there will ordinarily exist an output image g'(x,y) that is in better agreement with the noisy data than the true image is. Consequently, for the case of noisy data, a "solution" is not found until E_0 is decreased to a level somewhat less than $E_{|\hat{F}|}$.

For the astronomy problem one has a nonnegativity constraint in the object domain. Furthermore, one can compute upper bounds on the support of the object in any of several ways [25]. The simplest way is to use a rectangle that is half the size, in each dimension, of the smallest rectangle that encloses the autocorrelation. If the actual support of the object is known a priori, then that should of course be used. Any other types of a priori information should be used during the iterations if available. The support constraint can in general be defined by a binary mask which is unity within the support and zero outside; then the computations of Eqs. (7.4-4), (7.4-7) and (7.4-19) can be performed arithmetically without the use of logic, which is advantageous when using array processors.

There are many ways to pick an initial input to the algorithm. Although claims have been made that the crude phase estimation method described in Section 7.3B(v) offers a superior starting point [34], others have found that random numbers do as well or better [42]. Having an initial input close to the true solution reduces the number

of iterations required and might help to avoid some stagnation problems. If one of the reconstruction methods described in Section 7.5 has yielded an image, then that image would be an appropriate starting input. One can either view the other reconstruction method as a means for supplying starting inputs for the iterative transform algorithm or view the iterative transform algorithm as a means for "cleaning up" images reconstructed by the other method. If no other initial estimate for the object is available, then one should use random numbers in the object domain or for the Fourier phase, giving an unbiased start to the algorithm. In the object domain, a convenient starting guess, $g_0(x)$, can be formed by filling the support mask with random numbers. Another method [29] is to threshold the autocorrelation (at say, 0.005 its maximum value), demagnify that by a factor of two in each dimension by discarding every other row and every other column, and finally fill the resulting shape with random numbers. (Note that this shape does not necessarily contain the support of the object [25].)

The algorithm can be made to converge faster and avoid a stagnation problem [see Section 7.4C(iii)] if the support mask is chosen to be somewhat smaller than the correct support for the first cycle or two of iterations. Since it is the incorrect support, the smaller support mask is inconsistent with Fourier modulus and stagnation will eventually occur when it is used. Nevertheless, the smaller support mask helps to force most of the energy of the output, $g'(x,y)$, into a confined region in fewer iterations. After this has happened the support mask should be enlarged to the correct support constraint for the object. This enlargement of the mask could be done in more than one step if desired. When the algorithm has nearly finished reconstructing the object, it is often beneficial to make the support mask even larger than the correct support for the object. This helps to ensure that no parts of the object are being inadvertently truncated by the support constraint. The progression

from a smaller support mask to a larger one also helps to avoid having edges of the output image biased toward falling right at the edges of the support mask.

Recall from Section 7.4A that the heart of the algorithm consists of several cycles of iterations, where one cycle consists of K_1 iterations of the hybrid input-output algorithm [Eqs. (7.4-1), (7.4-2), (7.4-3) and (7.4-19)] followed by K_2 iterations of the error-reduction algorithm [Eqs. (7.4-1)-(7.4-4)]. Our experience has shown that values of K_1 from 20 to 100, of K_2 from 5 to 10, and of the feedback parameter β from 0.5 to 1.0 (use, say, 0.7) work well.

The discrete Fourier transforms are computed using the Fast Fourier Transform (FFT) algorithm. The sampling in the Fourier domain should be fine enough to ensure that the object domain array size is at least twice the width and height of the object itself, which is equivalent to achieving the Nyquist sampling rate for $|F(u,v)|^2$.

A straightforward method to evaluate Eq. (7.4-2) is to compute the phase from the real and imaginary parts of $G_k(u,v)$, then combine it with $|F(u,v)|$ to form $G'_k(u,v)$, and finally compute the real and imaginary parts of $G'_k(u,v)$ (which are required by the FFT) from its modulus and phase. Alternatively, one can employ

$$G'_k(u,v) = G_k(u,v)\ |F(u,v)|\ /\ [|G_k(u,v)| + \delta] \qquad (7.4\text{-}21)$$

where δ is a very small number used to prevent overflow problems in the rare event that $|G_k(u,v)| = 0$. (For some computers one can use $\delta = 0$ with no ill effects.)

If all goes well, the iterative transform algorithm will converge to a solution after a small number of cycles of iterations. If there are multiple solutions, the iterative transform algorithm is capable of finding any one of them, depending on the starting input [43, 44].

Confidence that the solution is the one and only true solution can be increased by performing two or more trials of the algorithm, each time using different random numbers for the initial input.

In some cases the iterative transform algorithm will stagnate before reaching a solution. The algorithm can be considered to have stagnated if the error E_0 is greater than $E_{|\hat{F}|}$ and has failed to decrease after three additional cycles. While some objects can be reconstructed very easily, requiring only one or two cycles, other more difficult objects can require many cycles comprising well over a thousand iterations. Consequently, one should not too readily jump to the conclusion that the algorithm has stagnated. It often occurs that very slow progress is made for many iterations, but then the algorithm suddenly finds its way and rapid progress is made in just a few iterations.

If the iterative transform algorithm does stagnate, then one can start over with a different set of random numbers for the initial input; alternatively one can try other simple alterations of the algorithm. One can use a larger value for β, say 1.2, causing larger changes to be made; but this should not be carried out for too many iterations since it causes the algorithm to become unstable. Use of a different support constraint mask may also be helpful. For certain special modes of stagnation there have been devised special methods for overcoming the stagnation, as described below.

C Methods for Overcoming Stagnation

Three particular modes of stagnation have been identified: those characterized by (1) simultaneous twin images, (2) stripes and (3) truncated support. Methods have been devised to jump each of these hurdles. They are briefly described in this section. further details can be found in Reference 45. (See Chapter 8 for another view of stagnation problems.)

(i) Simultaneous Twin Images

Since both $f(x,y)$ and its twin image $f(-x,-y)$ have the same Fourier modulus, the iterative transform algorithm is equally likely to reconstruct either one. When the support of $f(x,y)$ is symmetric, the algorithm often outputs a partially reconstructed image having features of both $f(x,y)$ and $f(-x,-y)$ as illustrated in Figure 7.4-2. In many cases the algorithm stagnates with such an output image. (It is not a solution because $E_0 > E_{|\hat{F}_1|}$.) If the algorithm has stagnated on an output image having a strong centro-symmetric component, then one would suspect this mode of stagnation. A method for circumventing this stagnation problem is to use a reduced-area temporary support constraint for a few iterations. The temporary support should be highly asymmetric and should include one or two edges of the object's support but not the opposite edge. After a few iterations with the reduced-area support constraint, the symmetry in the output image may be sufficiently broken so that upon further iterations using the correct support constraint the iterative transform algorithm can converge to either $f(x,y)$ or its twin.

For example, to overcome the simultaneous twin images stagnation problem exhibited in Figure 7.4-2(c), a triangular support covering the lower right half of the true square support was used for 10 error reduction iterations. When further iterations of the algorithm were performed with the correct square support, it converged to the solution.

(ii) Stripes

The algorithm frequently stagnates in a mode in which the output image is a faithful representation of the object but with a pattern of stripes superimposed. This can be recognized by virtue of the fact that the stripes extend (although with less contrast) beyond the support of the object. The near-sinusoidal nature of the stripes is an indication that the phase errors occur in a symmetric pair of

fairly-well-defined areas in the Fourier domain. It has been observed that, if the iterative transform algorithm is applied more than once to the same Fourier modulus data but with different random initial inputs, then the stripes (if there are any) of each of the reconstructed images tend to have different frequencies and orientations [46] which means that the errors occur in different areas of the Fourier domain. Figures 7.4-3(a) to 7.4-3(d) show an example of this.

One method for solving this stagnation problem is the voting method [45]. The iterative transform algorithm is applied three times, each time with a different random initial input to produce three different striped images, such as shown in Figures 7.4-3(b) to 7.4-3(d). (If one of them has no stripes, then of course, the problem is solved.) By cross-correlations the relative translations and orientations of the three images are determined and corrected so that they are made to be in perfect registration. The Fourier transforms of the three images are compared. At each point the two complex Fourier values that are closest are averaged and the third (assumed to have a phase error) is discarded. The resulting composite Fourier function is inverse transformed to produce the corresponding output image. Then further iterations of the iterative transform algorithm are performed. For example, Figure 7.4-3(e) shows the output of the voting method applied to the images shown in Figures 7.4-3(b) to 7.4-3(d), and Figure 7.4-3(f) shows the result of further iterations.

A second method for overcoming stagnation with stripes is the patching method [45]. It requires that only two striped images be produced and made to be in registration. For each of these images the area of the support of the object is zeroed out, leaving just the stripes in the area outside the object's support. These stripes are Fourier transformed, and the resulting Fourier modulus arrays are used to determine what areas in the Fourier domains have phase

errors associated with the stripes. It is assumed that those areas are different for each of two striped images. A new composite Fourier transform is set equal to the Fourier transform of the first striped image where it is not affected by the phase errors and to the Fourier transform of the second striped image where the first is affected by the phase errors. Both methods of overcoming the problem of stagnation with striped images have been shown to be effective [45].

(iii) Truncated Support

Since $f(x - x_o, y - y_o)$ has the same Fourier modulus as $f(x,y)$, the location of the object's support is arbitrary. Frequently the image partially reconstructed by the algorithm is not in perfect registration with the support constraint. Then enforcing the support constraint causes an inadvertent truncation of part of the desired image, causing the algorithm to stagnate. In addition to the enlarging support method described in Section 7.4B, a method of combatting this stagnation problem is to dynamically translate either the support constraint or the image.

The amount of translation to be used can be determined as follows. Compute the total energy of the output image, $g'_k(x,y)$, (i.e., square and sum) over the area of the support constraint for the current position of the support constraint and for the support constraint translated by one or two pixels in every direction. The support constraint should be translated to the position for which the energy is maximized. Alternatively, compute the cross-correlation of $g'_k(x,y)$ with the support mask, and translate according to the peak of the cross-correlation. This can be done occassionally or at every iteration. This method would be particularly effective if, just prior to performing it, a support constraint larger than the usual support were used for a few iterations; that would give the truncated part of the image a chance to establish itself.

D Iterative Reconstruction Example

In this section, the results of a computer experiment, in which stellar speckle interferometry [47] was realistically simulated -- including the effects of atmospheric turbulence and photon noise -- and reconstruction from noisy Fourier modulus data by the iterative transform algorithm, are shown [46]. (See Section 7.5A for a description of stellar speckle interferometry.)

An undegraded object, a low-pass filtered, digitized photograph of a satellite, shown in Figure 7.4-4(a), was convolved with 156 different point spread functions to produce 156 different blurred images. Each of the point spread functions represented a different realization of the effects of the turbulent atmosphere. The blurred images were then subjected to a Poisson noise process to simulate the effects of photon noise. For this example, there were approximately 300,000 photons per degraded image (or on the order of 100 photons per pixel over the extent of the object), which is realistic for objects of this type when imaged through a telescope of diameter 1.6 m. Two of the resulting 156 degraded images are shown in Figures 7.4-4(b) and 7.4-4(c). The degraded images were then processed by Labeyrie's [47] method as modified by Goodman and Belsher [48]. The estimate of the modulus of the Fourier transform of the object is given by [46]

$$F(u,v) = W(u,v)\left[\frac{\sum_{m=1}^{M} |I_m(u,v)|^2 - N_p}{\sum_{m=M+1}^{2M} |T_m(u,v)|^2}\right]^{1/2}, \qquad (7.4\text{-}22)$$

where $I_m(u,v)$ is the Fourier transform of the m^{th} degraded image; N_p is the total number of photons detected (it is subtracted in order to compensate for a noise bias term that arises in the power

spectrum due to photon noise [48]); $T_m(u,v)$ is the Fourier transform of the m^{th} point-spread function [to provide compensation for the modulation transfer function (MTF) of the speckle interferometry process]; and the weighting factor $W(u,v)$ is the MTF due to the telescope aperture. In practice, the denominator of this expression would be obtained by making measurements on a reference star through an atmosphere having the same satistics as that which blurred the images or by using a model of the effects of atmospheric turbulence. The term $W(u,v)$ was included in order to restore the natural MTF due to the telescope aperture which was removed by the denominator of Eq. (7.4-22). Figure 7.4-4(d) shows the resulting Fourier modulus estimate.

The object was reconstructed (or equivalently, the Fourier phase was retrieved) using the hybrid input-output algorithm alternately with the error-reduction algorithm. The result, shown in Figure 7.4-4(e), agrees very well with the original object shown in Figure 7.4-4(a), despite the noise present in the Fourier modulus data. Good reconstructed images were also obtained when only one-tenth as many photons were assumed to be available [46].

For this reconstruction example, Figure 7.4-5 shows E_o versus the number of iterations and Figure 7.4-6 shows the output image at various stages. The starting input used, shown in Figure 7.4-6(b) was the randomized, demagnified autocorrelation described in Section 7.4B, using a threshold value of 0.0004 times the peak of the autocorrelation. For the first 10 iterations, the error-reduction algorithm was used and the mask defining the support constraint was chosen to be the region over which the autocorrelation function, spatially demagnified by a factor of 2, exceeded 0.0004 of its maximum value, providing a fairly tight diameter constraint [it is just the support of Figure 7.4-6(b), which is shown in Figure 7.4-6(a)]. For iterations 11 to 20, the error-reduction algorithm was used, and the mask for these and the remaining iterations was

chosen to be a square of length 64 pixels, which is larger than the actual object extent of about 60 x 40 pixels (imbedded in an array of 128 x 128 pixels). The error E_0 decreased rapidly at first, and the visual appearance of the output image improved substantially from the first iteration, shown in Figure 7.4-6(c), to the fifth iteration, shown in Figure 7.4-6(d). The error decreased suddenly at the tenth iteration since some positive-valued points that were inside the second mask but outside the first mask were no longer counted as contributing to E_0. By the twentieth iteration, the error-reduction algorithm was converging very slowly, and the output image, shown in Figure 7.4-6(e), is only slightly improved over that in Figure 7.4-6(d). For iterations 21 to 60, the hybrid input-output algorithm, with β equal to 1, was used. At first E_0 increased sharply, although, as shown in Figure 7.4-6(f), the output image at iteration 25 appeared no worse than at iteration 20; then E_0 decreased fairly rapidly until stagnating at $E_0 \approx 0.05$ at about iteration 55. Output images at iterations 35, 45, and 55 are shown in Figures 7.4-6(g), (h) and (i), respectively. For iterations 61 to 70, the error-reduction algorithm was used, for which E_0 dropped suddenly from 0.05 to 0.02, although the visual appearance of the reconstructed image remained the same as for iteration number 55.

This final value of E_0 is somewhat less than 0.03, the normalized rms error of the Fourier modulus estimate itself.

Reconstruction experiments do not always proceed as smoothly as the one described above. When stagnation occurs, then the methods described in Section 7.4B and 7.4C should be employed.

7.5 Solutions Specific to Measurement Techniques

We have assumed so far in this chapter that only the modulus of the Fourier transform, $|F(u,v)|$, of the object intensity, $f(x,y)$, is known and have combined this data with the a priori knowledge that

the object is nonnegative and of finite extent to achieve a reconstruction of the object. Astronomical observing techniques based on either speckle (image plane) or Michelson (pupil plane) interferometry can yield data in addition to the modulus of the Fourier transform: for example, speckle interferometry gives, in addition to a diffraction-limited estimate of the modulus, a low resolution image of the object, i.e., an estimate of both the modulus and phase of the transform at low spatial frequencies. This additional data may in some cases be sufficient to determine the object intensity without any ambiguity. Since the different methods are specific to the observing technique, they are described separately under the headings of Speckle Interferometry and Michelson Interferometry.

A. Object Recovery from Speckle Interferometry Data

(i) Basic Speckle Technique

Speckle interferometry was invented by A. Labeyrie [47] in 1970 as a means of obtaining diffraction-limited resolution (but not images) from large telescopes, despite the presence of atmospheric turbulence or "seeing". The method is illustrated in Figure 7.5-1 for an unresolved star, binary stars of two separations, and a resolved star. A large number of short-exposure (~10 ms) photographs are taken, each through a different realization of the atmosphere, typical examples being shown in row B (see also Figure 7.5-2). For a binary star, each component produces an identical speckle pattern and a "double-speckle" effect may be visible in each short-exposure image in favorable circumstances. The optical diffraction pattern, or squared modulus of the Fourier transform, of each photograph is shown in row C -- the signal-to-noise ratio is low for a single record and may be improved by adding many such diffraction patterns (row D). The unresolved object has a diffraction halo of relatively large spatial extent, the binaries give fringes of period inversely

proportional to their separation, and the resolved object gives a diffraction halo whose diameter is inversely proportional to the diameter of the object. By taking a further Fourier transform of each ensemble-averaged diffraction pattern, we obtain the average spatial autocorrelation of the diffraction-limited images of each object (row E).

In mathematical terms, one estimates either the average spatial autocorrelation, $\langle g(x,y) \star\!\!\star g(x,y) \rangle$, or the power[+] spectrum $\langle |G(u,v)|^2 \rangle$, of the image intensity $g(x,y)$, where $\langle\ \rangle$ denotes the ensemble average, from a finite number of frames. It can be shown that the image power spectrum is related to that of the object, $|F(u,v)|^2$, through the speckle transfer function $\langle |T(u,v)|^2 \rangle$:

$$\langle |G(u,v)|^2 \rangle = |F(u,v)|^2 \langle |T(u,v)|^2 \rangle, \qquad (7.5\text{-}1)$$

where

$$\langle |T(u,v)|^2 \rangle \approx |\langle T(u,v) \rangle|^2 + T_D(u,v)/N_s\ , \qquad (7.5\text{-}2)$$

and all spatial frequencies are referred to those of the image (formed by a telescope of focal length z using light of wavelength λ). In Eq. (7.5-2), $\langle T(u,v) \rangle$ is the average (seeing-limited) transfer function, $T_D(u,v)$ is the <u>diffraction-limited</u> transfer function of the telescope and N_s is the average number of speckles per frame [$\approx 2.3(D/r_0)^2$], where r_0 is the atmospheric seeing parameter).

The key feature of speckle interferometry is that the speckle transfer function $\langle |T(u,v)|^2 \rangle$ is nonzero up to the diffraction

[+]Strictly speaking this should be called the average <u>energy</u> spectrum [3, p. 260].

limit of the telescope, as given by Eq. (7.5-2), allowing the object power spectrum $|F(u,v)|^2$, or equivalently the object autocorrelation $f(x,y) \star f(x,y)$, to be estimated up to the diffraction limit of the telescope. Note that the basic technique does not give the object intensity itself: to address this problem, one must use one of the phase retrieval algorithms described earlier or one of the speckle imaging techniques described below.

(ii) Shift-and-Add Method

The first speckle imaging technique to be suggested used the notion that the individual speckles in a short-exposure image resemble diffraction-limited images of the object. Figure 7.5-2 compares short-exposure images of a point source and a resolvable star and there is a clear qualitative difference between the speckle structure in the two cases. By averaging the centroided versions of selected bright speckles, Lynds et al [49] were able to reconstruct a map of the red giant α-Orionis. The procedure was made a little more systematic and generalized by Bates and Cady [50], who suggested that the brightest speckles in the image be shifted to the origin and the entire images added (hence "shift-and-add") allowing for larger images to be formed.

There are two problems with this method that make it unlikely to be of practical significance for most objects. First, to be able to implement the method, it is necessary to identify individual speckles and this can only be done if there are more than approximately 10^5 detected photons per frame, implying an object brighter than about 5th magnitude in the visible. Second, each speckle is obviously not an image of the object and there is no linear transfer function for this technique, making it difficult to reconstruct reliable quantitative object maps. The Bates and Cady extension can give good results if the image contains a very bright unresolved component [51].

(iii) Cross-Power Spectrum Method

This method, first suggested by Knox and Thompson [52], makes use of the fact the phase difference between two values of $F(u,v)$, separated by $(\Delta u,\Delta v) < r_0/\lambda z$ in spatial frequency, is present in the original speckle data $g(x,y)$ and can be extracted by computing the cross-power spectrum $\langle G(u,v)G^*(u + \Delta u, v + \Delta v)\rangle$. A straightforward analysis shows that

$$\text{phase}\left\{\langle G(u,v)G^*(u + \Delta u,\ v + \Delta v)\rangle\right\}$$
$$= \text{phase}\left\{F(u,v)\right\} - \text{phase}\left\{F(u + \Delta u,\ v + \Delta v)\right\}$$
$$+ \text{phase}\left\{\langle T(u,v)T^*(u + \Delta u,\ v + \Delta v)\rangle\right\} . \qquad (7.5\text{-}3)$$

The phase of a complex quantity only has meaning if its modulus is nonzero, so that from Eq. (7.5-3) there are two conditions to be met if phase differences of $F(u,v)$ are to be computed from the phase of the cross-spectrum:

(i) $\left|\langle T(u,v)T^*(u + \Delta u,\ v + \Delta v)\rangle\right| \neq 0,$

(ii) $\text{phase}\left\{\langle T(u,v)T^*(u + \Delta u,\ v + \Delta v)\rangle\right\}$ either known or zero.

If we assume that the complex amplitude in the telescope pupil has complex Gaussian statistics and that its autocorrelation function has a Gaussian shape, i.e., the "seeing" transfer function $T_s(u,v)$ is Gaussian, then it can be shown [3, p. 303] that

$$\langle T(u,v)T^*(u + \Delta u,\ v + \Delta v)\rangle \approx \langle |T(u,v)|^2\rangle\ |T_s(\Delta u/2,\ \Delta v/2)|^2 . \qquad (7.5\text{-}4)$$

That is, the Knox-Thompson transfer function is simply the product of the normal speckle transfer function (Eq. 7.5-2) at (u,v) and the squared modulus of the "seeing" transfer function at $(\Delta u/2,\Delta v/2)$. It follows that the modulus of the Knox-Thompson transfer function is nonzero up to the diffraction limit of the telescope provided that $\Delta u/2$ and $\Delta v/2$ are less than $(r_0/\lambda z)$, the width of the "seeing" transfer function. It also follows that

$$\text{phase}\left\{\langle T(u,v)T^*(u+\Delta u, v+\Delta v)\rangle\right\} \simeq 0.$$

The problem of finding the phase of $F(u,v)$ given the phase differences for frequencies separated by $(\Delta u, \Delta v)$ is identical to that of finding a wavefront from sheared interferograms [53]. Simple "bootstrapping" of the phase differences, starting at the origin of the spatial frequency plane, leads to cumulative errors at higher frequencies and in two dimensions yields phase values at a given frequency that depend on the path taken in the frequency plane: averaging of these values is not the best way to deal with this problem. A least-squares solution is straightforward to implement using an iterative technique to solve the usual matrix equation (the matrix that relates the phase differences to the phase values is sparse) [54].

The assumptions leading to Eq. (7.5-4) are not particularly realistic and it remains to be seen how sensitive this technique is to the atmospheric statistics.

(iv) Triple Correlation Method

The average triple correlation $\langle g^{(3)}(\underline{x}_1,\underline{x}_2)\rangle$ of the image intensity $g(\underline{x})$, or equivalently its Fourier transform, the bispectrum $\langle G^{(3)}(\underline{u}_1,\underline{u}_2)\rangle$, contains information on the phase of $G(\underline{u})$ and hence on the phase of the object transform [55, 56]. Here we use the notation $\underline{x} = (x,y)$ and $\underline{u} = (u,v)$. The triple correlation and bispectrum are, for real function $g(\underline{x})$, defined by:

$$g^{(3)}(\underline{x}_1,\underline{x}_2) = \int\int_{-\infty}^{\infty} g(\underline{x})\, g(\underline{x}+\underline{x}_1)\, g(\underline{x}+\underline{x}_2)\, d\underline{x} \qquad (7.5\text{-}5)$$

and

$$G^{(3)}(\underline{u}_1,\underline{u}_2) = G(\underline{u}_1)\, G(\underline{u}_2)\, G(-\underline{u}_1-\underline{u}_2) \quad . \qquad (7.5\text{-}6)$$

The average bispectrum of the image is related to that of the object through a bispectrum transfer function $\langle T^{(3)}(\underline{u}_1,\underline{u}_2)\rangle$,

$$\langle G^{(3)}(\underline{u}_1,\underline{u}_2)\rangle = F^{(3)}(\underline{u}_1,\underline{u}_2)\, \langle T^{(3)}(\underline{u}_1,\underline{u}_2)\rangle \qquad (7.5\text{-}7)$$

If it is assumed that complex amplitude in the telescope pupil has Gaussian statistics, then it can be shown [55] that the bispectrum transfer function is real and nonzero over the diffraction-limited portion of the $(\underline{u}_1, \underline{u}_2)$ volume: it follows that the phase of the image bispectrum equals that of the object bispectrum.

To see how the phase of the object spectrum (which is Hermitian) can be calculated recursively from that of the object bispectrum, consider a one-dimensional sampled bispectrum $F^{(3)}_{\ell,m}$:

$$F^{(3)}_{\ell,m} = F_\ell \, F_m \, F_{-\ell-m}. \tag{7.5-8}$$

Denoting the phase of the object spectrum by ψ_n and that of the bispectrum by $\beta_{\ell,m}$, the phase factors satisfy the equation

$$\exp [i\psi_n] = \exp [i(\psi_m + \psi_{n-m} - \beta_{n-m,m})] \tag{7.5-9}$$

Where the substitution $n = \ell + m$ has been made. Note that both ψ_0 and ψ_1 may be set equal to zero as the object is real and its absolute position is not of interest.

Now consider Eq. (7.5-9) with $m = 1$ and $n = 2, 3, 4, \ldots$:

$$\exp [i\psi_n] = \exp [i(\psi_1 + \psi_{n-1} - \beta_{n-1,1})], \ (\psi_1 = 0). \tag{7.5-10}$$

Clearly, $\psi_2, \psi_3, \ldots$ can be found by simple recursion with $n = 2, 3 \ldots$. In fact, the process can be repeated for different values of m, to yield a more accurate estimate of the phase at each spatial frequency. The method is in some ways similar to, but more general than, the cross-power spectrum technique described in (iii).

The triple correlation method is illustrated in Figure 7.5-3 for a one-dimensional projection of the double star ω-Leonis. For the more interesting case of two-dimensional objects, the reconstruction of the object spectrum from the bispectrum is computationally quite reasonable -- the problem lies in computing the average bispectrum,

which is a four-dimensional function. Nevertheless, the triple correlation method appears to be the most promising technique of speckle imaging at the present time.

(v) Exponential Filter Method

For discrete objects, the object spectrum can be written as a polynomial or z-transform, as discussed in Chapter 6.1. For modulus-only reconstruction in one dimension, the possibility of ambiguities arises because the zeros of the power spectrum of the object include both the zeros of the original object and their inverses: in the one-dimensional case it is not in general possible to distinguish the correct zeros from their inverses, given only the Fourier modulus. The cross-spectrum method eliminates this ambiguity by rotating the "inverse" zeros by a known amount [57], thus enabling them to be distinguished from the correct ones even in the one-dimensional case. The triple correlation method also allows the zeros of the object spectrum to be identified correctly from the zeros of the triple correlation [56]. The exponential filter method provides a third way of distinguishing the correct zeros from their inverses: in this case the correct zeros are moved out radially in the complex plane by a known amount.

In this method, one records the instantaneous image intensity $g(x,y)$ in the usual way and computes the average power spectrum $\langle|G(u,v)|^2\rangle$. A second power spectrum, $\langle|G'(u,v)|^2\rangle$ of $g(x,y)\ p(x,y)$, where $p(x,y)$ is a special modifying function, is also computed. Walker [58] showed that if the modifying function has the property,

$$p(x_1 + x_2,\ y_1 + y_2) = p(x_1,y_1)\ p(x_2,y_2) \quad ,$$

then

$$\langle|G'(u,v)|^2\rangle = |F'(u,v)|^2\ \langle|T'(u,v)|^2\rangle \qquad (7.5\text{-}11)$$

where

$$F'(u,v) = F(u,v) * P(u,v)$$

and $\langle |T'(u,v)|^2 \rangle$ is a transfer function that can be measured by observing a point source. The exponential, $p(x,y) = \exp(-2\pi ax)$, has the required property. From the data we have thus extracted the power spectrum of the object, $|F(u,v)|^2$, and that of the exponentially-filtered object, $|F'(u,v)|^2$: these together, in principle, contain sufficient information to reconstruct the object intensity.

In practice, in the two-dimensional case, an iterative algorithm similar to that described in Section 7.4 is used to obtain the object intensity $f(x,y)$ from these two power spectra, which are both used alternately as Fourier constraints [59]. Figure 7.5-4 shows an example of reconstructions obtained in a computer simulation using this algorithm.

(vi) Phase Averaging Method

The instantaneous image intensity $g(x,y)$ is related to the object intensity $f(x,y)$ by the usual convolution formula, which in Fourier space becomes a product:

$$G(u,v) = F(u,v)\ T(u,v)\ , \qquad (7.5\text{-}12)$$

where $T(u,v)$ is the instantaneous transfer function of the atmosphere/telescope combination. Taking the ensemble average of the logarithm of Eq. (7.5-12) and equating imaginary parts, we obtain

$$\langle \text{phase}\ \{G(u,v)\} \rangle = \text{phase}\ \{F(u,v)\} + \langle \text{phase}\ \{T(u,v)\} \rangle + 2\pi q, \qquad (7.5\text{-}13)$$

where the phase is the unwrapped value in the interval $-\infty$ to $+\infty$ and q is an integer. McGlamery [60] suggested that if $\langle \text{phase}\ T(u,v) \rangle$ is known (or zero), then the phase of the object transform could be found from the average phase of the image transform. Using the Central Limit Theorem, it is possible to show that, for spatial

frequencies $r_0/\lambda z < (u,v) < (D - r_0)/\lambda z$, the quantity $T(u,v)$ is a circular complex Gaussian process, with the result that $\langle \text{phase} \{T(u,v)\} \rangle \approx 0$.

The real problem with this phase averaging method is determining, using continuity of $G(u,v)$, the unwrapped phase (the phase in the interval $-\infty$ to $+\infty$) from the phase of $G(u,v)$, which is necessarily in the interval $-\pi$ to $+\pi$. This unwrapping has to be done before the averaging process. When the modulus $|G(u,v)|$ is small, the rms absolute error σ in the unwrapped phase is given approximately by,

$$\sigma \approx \frac{1}{\sqrt{2\bar{N}}\,|G(u,v)|} \qquad (7.5\text{-}14)$$

where $\bar{N}$ is the average number of detected photons per frame and $G(u,v)$ is normalized to unity at zero spatial frequency. In addition to having a low <u>average</u> value of its modulus [because of the form of the speckle transfer function, Eq. (7.5-2)], the continuous function $G(u,v)$, since it is the Fourier transform of a speckle pattern, also contains many points whose modulus is zero, which can be expected to cause problems in phase unwrapping algorithms [61].

Mertz [62] has suggested that it should be possible to use continuity in <u>time</u>, as well as spatial frequency, to help unwrap the phase and such a proposal may be easy to implement using the time-ordered photon event detectors that are becoming available.

(vii) <u>Non-Redundant Aperture Methods</u>

As is well known, the Fourier transform of any intensity point spread function is equal to the autocorrelation of the complex amplitude distribution over the pupil. A given spatial frequency component is therefore the sum (or integral) of products of values of the complex amplitude at all combinations of two points in the pupil separated by a fixed amount. If a special mask of sub-apertures is used in the pupil, such that no vector separation

occurs more than once (i.e., a non-redundant array of sub-apertures), then each spatial frequency component of the point spread function equals the product of two particular values of the complex amplitude in the pupil and hence yields the phase difference between the wavefront at these two points. Thus, the Fourier transform of the image intensity, when the imaging aperture is a non-redundant array of sub-apertures, has a simple, direct interpretation in terms of the wavefront in the telescope aperture, and this has been the basis of several possible methods of diffraction-limited imaging from (speckle-free) image plane data. However, in all cases each frame of data has to have an excellent signal-to-noise ratio so that an accurate phase difference can be found.

Three methods of utilizing short-exposure images taken through non-redundant apertures have been put forward. Rhodes and Goodman [63] suggested using two or more masks to form two or more short-exposure images with the same atmospheric distortion -- this procedure could be used to eliminate telescope aberrations as well. Brown [64] suggested that even a single image from a single mask could be processed using a suitable "sharpness" criterion for the reconstructed object. Finally, there is the possibility of phase averaging analogous to that described in Section (vi) above.

A variation on the use of a non-redundant aperture has been proposed by Greenaway [65]. The basic idea is to reconfigure a redundant aperture into a nearly non-redundant one using a suitable interferometer and then applying phase closure techniques. This allows the full aperture to be utilized, but it is not clear how one constructs the optical system required to reconfigure the aperture.

(viii) Phase Diversity Method

As was seen in Section 7.5A(v), the availability of more than one type of measurement opens up additional possibilities for reconstructing an image. A class of such techniques are those that

employ phase diversity in the pupil of the telescope [66]. Two images are collected simultaneously: the usual aberrated image with optical transfer function T(u,v) and Fourier transform

$$G(u,v) = F(u,v)\ T(u,v), \tag{7.5-15}$$

and a "phase-diversity" image with optical transfer function $T_d(u,v)$ and Fourier transform

$$G_d(u,v) = F(u,v)\ T_d(u,v). \tag{7.5-16}$$

The phase-diversity transfer function, $T_d(u,v)$, includes a known, purposely-induced phase aberration in addition to the same unknown atmospherically-induced aberrations included in T(u,v). For this problem, both the object and the atmospheric aberrations are unknown.

To solve this problem, Gonsalves [66] introduced the error metric

$$E = \iint \left[|G(u,v) - \hat{F}(u,v)\hat{T}(u,v)|^2 + |G_d(u,v) - \hat{F}(u,v)\hat{T}_d(u,v)|^2 \right] du\ dv \tag{7.5-17}$$

(where $\hat{X}$ is an estimate of X). By the method of calculus of variations, the function $\hat{F}(u,v)$ that minimizes E can be determined and inserted in Eq. (7.5-17) to yield an expression for the error metric that is independent of $\hat{F}(u,v)$:

$$E = \iint \left[\frac{|G(u,v)\hat{T}_d(u,v) - G_d(u,v)\hat{T}(u,v)|^2}{|\hat{T}(u,v)|^2 + |\hat{T}_d(u,v)|^2} \right] du\ dv \tag{7.5-18}$$

This error metric depends only on the estimates of the transfer functions which in turn depend only on the estimate of the phase aberrations. One can iteratively search for the estimate of the phase aberrations (usually parameterized by, say, the coefficients of a Zernike polynomial expansion) that minimize E, by the Fletcher-Powell gradient search method for example. Having found the aberrations, T(u,v) can be computed and F(u,v) can be determined from Eq. (7.5-15) or, better, by Wiener filtering.

A problem with this method is that, employing only a single pair of frames of data, it works well only for bright objects.

B Object Recovery from Michelson Interferometry Data

(i) Pupil Plane Interferometry

In pupil plane interferometry, the spatial coherence function of starlight is measured by simple two-beam interference either directly in the pupil of a large telescope or using two (or more) separated telescopes in a long baseline arrangement.

The spatial coherence function, $C(\Delta\xi,\Delta\eta)$, is defined by

$$C(\Delta\xi,\Delta\eta) \equiv \langle V(\xi,\eta,t)\, V^*(\xi + \Delta\xi,\, \eta + \Delta\eta,t)\rangle \,, \qquad (7.5\text{-}19)$$

where $V(\xi,\eta,t)$ is the analytic signal representation of the field in the (ξ,η) plane. The real part of $C(\Delta\xi,\Delta\eta)$ can be measured by interfering portions of the wavefront located at (ξ,η) and $(\xi + \Delta\xi,\eta + \Delta\eta)$ and measuring the resultant time-average[+] intensity:

$$\langle |V(\xi,\eta,t) + V(\xi + \Delta\xi,\, \eta + \Delta\eta,t)|^2 \rangle$$
$$= I(\xi,\eta) + I(\xi + \Delta\xi,\, \eta + \Delta\eta) + 2\,\mathrm{Re}[C(\Delta\xi,\Delta\eta)]. \qquad (7.5\text{-}20)$$

The imaginary part is found by retarding one wavefront by $\pi/2$ prior to interference. In practice, the two wavefronts are either combined through a beamsplitter arrangement to yield both the real and imaginary parts of $C(\Delta\xi,\Delta\eta)$ simultaneously, or at an angle so that the fringe positions gives the phase (if it is possible to follow individual fringes).

[+]Or ensemble average, the distinction is unimportant for ergodic light sources such as stars.

According to the van-Cittert-Zernike theorem, the coherence function $C(\Delta\xi,\Delta\eta)$ is proportional to the Fourier transform $F(u,v)$ of the object intensity $f(x,y)$: if we take the object coordinates referred to image space, the spatial frequencies (u,v) of the object transform are related to separation in the pupil plane $(\Delta\xi,\Delta\eta)$ given by

$$u = \Delta\xi/\lambda z, \qquad v = \Delta\eta/\lambda z,$$

where z is the focal length of the telescope[++].

Although historically this method was first proposed and demonstrated by Fizeau and Stephan, it is normally known as Michelson interferometry, following Michelson's measurements of the satellites of Jupiter and the red giant, α-Orionis, the latter using a specially constructed two-beam interferometer mounted on the 100-inch Mt. Wilson telescope. The main attraction of the implementation of Michelson interferometry using two small telescopes is that the baseline, which in terms of resolution is equivalent to the diameter of a single aperture telescope, can potentially be very large, perhaps 1 km or more. It is also relatively easily adapted to use in space. The major disadvantages of the two-telescope approach are the poor coverage of the (u,v) plane and the lack of reliable phase information.

(ii) Wavefront Rotational Shearing Methods

This method is equivalent to an array of Michelson interferometers working in parallel and is an alternative to speckle (image plane) methods for single large telescopes. Mertz [67] was

[++]It is customary in astronomy to use angular frequencies $\Delta\xi/\lambda$ and $\Delta\eta/\lambda$.

the first to point out that a rotational shearing interferometer can be used to measure the two-dimensional Fourier transform of any incoherent source distribution with a variable magnification that depends on the amount of rotational shear. This interferometer has been adapted by Roddier [68] for astronomical observation.

In a rotational shearing interferometer, the time-varying complex amplitude in the telescope pupil, $V_1(\xi,n,t)$, is divided by a beamsplitter into two beams, one of which is rotated by an angle θ with respect to the other before recombination at the same or another beamsplitter. Variable magnification is achieved by varying θ, but to see the principle of operation we shall let $\theta = 180^\circ$. The output intensity $I_2(\xi,n)$ of the interferometer in this case is

$$I_2(\xi,n) = \langle |V_1(\xi,n,t) + V_1(-\xi,-n,t)|^2 \rangle$$

$$= 2\, I_1(\xi,n) + 2\, \mathrm{Re}[C(2\xi,2n)] \qquad (7.5\text{-}21)$$

for a uniform pupil intensity. The imaginary part of the spatial coherence could be found separately by phase-shifting one beam prior to recombination or, more simply, by adding tilt fringes to form a side-band hologram [68].

In applying this method to imaging through turbulence, it is important to note that the fringe <u>visibility</u>, from which the spatial coherence is found, remains constant from frame-to-frame and the turbulence affects only the detailed form of the fringes. The major disadvantage of the technique is that each frame must have sufficient signal-to-noise ratio to enable the fringe visibility to be determined and this requires that the individual fringes are clearly defined. This is a significant drawback despite the fact that at high light levels pupil plane interferometers have a higher signal-to-noise ratio than the image plane technique of speckle interferometry.

(iii) Phase Closure

Phase retrieval in the two (or more) telescope arrangement originally proposed by Michelson is very much more complicated than for the single aperture case, as there are almost certainly discontinuities of the measurements in the pupil plane because the full aperture is sparsely filled. One approach is to combine simultaneously the outputs from a non-redundant coherent array and use the image plane (speckle) data as described in Section 7.5A(viii). Another technique, originally suggested and used in radio-astronomy [69], is to form simultaneously all possible pair correlations from N apertures using the Michelson technique to determine the Fourier phase, and then use the principle of phase closure to eliminate some or (given enough redundant measurements) all the (unknown) atmospheric phase disturbances.

Consider any three apertures and let the (constant) phase across each one due to atmospheric turbulence or other phase errors be α_1, α_2 and α_3. The measurements of instantaneous fringe positions for each aperture pair yields phase differences β_{12}, β_{23} and β_{13} which are related to the true object phase differences ψ_{12}, ψ_{23} and ψ_{13} by

$$\beta_{12} = \psi_{12} + \alpha_1 - \alpha_2$$

$$\beta_{23} = \psi_{23} + \alpha_2 - \alpha_3$$

$$\beta_{13} = \psi_{13} + \alpha_1 - \alpha_3.$$

By forming the sum

$$\beta_{12} + \beta_{23} - \beta_{13} = \psi_{12} + \psi_{23} - \psi_{13}$$

the atmospheric phase terms are eliminated. If one assumes that two object phases are known, then the third can be calculated. The

arbitrary choice of the first two phases simply affects the position of the reconstructed object on the sky and does not affect its morphology. Repeating this with a carefully chosen set of aperture spacings (which include redundant spacings) it is possible to determine the Fourier phase of the object at all spatial frequencies covered by the array. Even if sufficient redundant spacings are not present to ensure a complete reconstruction, the number of unknown phases is greatly reduced by the phase closure method.

7.6 CONCLUSIONS

The methods for reconstructing stellar objects f(x, y) from interferometric data in optical astronomy fall into two classes: those that require only the Fourier modulus $|F(u,v)|$ and those that use the additional, but limited, phase data that is available in certain observational techniques.

It appears that object reconstruction from the Fourier modulus alone is usually unique, although it is not difficult to construct specific counterexamples that do not yield a unique solution. The main problem is how to find this solution. The iterative transform algorithm described in Section 7.4 is the most computationally efficient and successful algorithm to date. Using array processors of modest cost, a complete iteration on a 128 x 128 data set need only take about 1 second and convergence to a solution can be obtained in a few minutes if stagnation problems are avoided (see Section 7.4C).

Particular observing techniques, such as speckle interferometry, can provide not only the Fourier modulus but also Fourier phase information. Clearly, if this additional information is available, it should be used in the reconstruction process. The most promising techniques for doing this at the moment are the cross-spectrum, triple correlation and exponential filter methods described in Sections 7.5A(iii)-(v) respectively.

The planned construction of several giant 10-20 m diameter ground-based optical telescopes in the near future should provide the stimulation for further advances in phase retrieval and object reconstruction in astronomy.

Acknowledgements

One author (JRF) would like to thank the U.S. Air Force Office of Scientific Research for its financial support under Contract F49620-82-K-0018.

REFERENCES

1. J.W. Goodman (1984). "Statistical Optics," Wiley, New York.

2. R.H.T. Bates (1982). "Astronomical Speckle Imaging," Physics Reports (Reviews Sect. of Phys. Lett.) 90, 203-297.

3. J.C. Dainty (1984). "Laser Speckle and Related Phenomena," (Springer Verlag, Heidelberg. 2nd edition), pp. 255-320.

4. J.C. Dainty (1984). "Progress in Image Reconstruction in Astronomy," Proc. SPIE, 492.

5. A. Walther (1963). "The Question of Phase Retrieval in Optics," Optica Acta 10, 41-49.

6. A.H. Greenaway (1977). "Proposal for Phase Recovery from a Single Intensity Distribution," Opt. Lett. 1, 10-12.

7. T.R. Crimmins and J.R. Fienup (1981). "Ambiguity of Phase Retrieval for Functions with Disconnected Support," J. Opt. Soc. Am. 71, 1026-1028.

8. T.R. Crimmins and J.R. Fienup (1983). "Uniqueness of Phase Retrieval for Functions with Sufficiently Disconnected Support," J. Opt. Soc. Am. 73, 218-221.

9. J.R. Fienup (1978). "Reconstruction of an Object from the Modulus of Its Fourier Transform," Opt. Lett. 3, 27-29.

10. P.J. Napier and R.H.T. Bates (1974). "Inferring Phase Information from Modulus Information in Two-Dimensional Aperture Synthesis," Astron. Astrophys. Suppl. 15, 427-430.

11. W. Lawton (1980). "A Numerical Algorithm for 2-D Wavefront Reconstruction from Intensity Measurements in a Single Plane," in 1980 International Optial Computing Conference, W.T. Rhodes, ed., Proc. SPIE 231, 94-98.

12. M. Nieto-Vesperinas (1980). "Dispersion Relations in Two-Dimensions: Application to the Phase Problem," Optik (Stuttgart) 56, 377-384.

13. I. Manolitsakis (1982). "Two-Dimensional Scattered Fields: A Description in Terms of the Zeros of Entire Functions," J. Math. Physics 23, 2291-2298.

14. R. Barakat and G. Newsam (1984). "Necessary Conditions for a Unique Solution to Two-Dimensional Phase Recovery," J. Math. Phys. 25, 3190-3193.

15. J.L.C. Sanz and T.S. Huang (1983). "Unique Reconstruction of a Band-Limited Multidimensional Signal from its Phase or Magnitude," J. Opt. Soc. Am. 73, 1446-1450.

16. Yu.M. Bruck and L.G. Sodin (1979). "On the Ambiguity of the Image Reconstruction Problem," Opt. Commun. 30, 304-308.

17. M.A. Fiddy, B.J. Brames, and J.C. Dainty (1983). "Enforcing Irreducibility for Phase Retrieval in Two Dimensions," Opt. Lett. 8, 96-98.

18. J.R. Fienup (1983). "Reconstruction of Objects Having Latent Reference Points," J. Opt. Soc. Am. 73, 1421-1426.

19. B.J. Brames (1985). "Uniqueness and Other Aspects of the Optical Phase Problem," Ph.D. Thesis, University of Rochester.

20. T.R. Crimmins (April 1986). "Phase Retrieval for Discrete Functions with Support Constraints: Summary," in OSA Topical Meeting on Signal Recovery and Synthesis II, Honolulu, Hawaii.

21. J.L.C. Sanz T.S. Huang, and F. Cukierman (1983). "Stability of Unique Fourier-Transform Phase Reconstruction," J. Opt. Soc. Am. 73, 1142-1145.

22. E.N. Leith and J. Upatnieks (1962). "Reconstructed Wavefronts and Communication Theory," J. Opt. Soc. Am. 52, 1123-1130.

23. J.W. Goodman (1970). "Analogy Between Holography and Interferometric Image Formation," J. Opt. Soc. Am. 60, 506-509.

24. C.Y.C. Liu and A.W. Lohmann (1973). "High Resolution Image Formation through the Turbulent Atmosphere," Opt. Commun. 8, 372-377.

25. J.R. Fienup, T.R. Crimmins, and W. Holsztynski (1982). "Reconstruction of the Support of an Object from the Support of Its Autocorrelation," J. Opt. Soc. Am. 72, 610-624.

26. G.H. Stout and L.H. Jensen (1968). "X-Ray Structure Determination" (Macmillan, London).

27. J.E. Baldwin and P.J. Warner (1978). "Phaseless Aperture Synthesis," Mon. Not. R. Astr. Soc. 182, 411-422.

28. B.R. Frieden and D.G. Currie (1976). "On Unfolding the Autocorrelation Function," J. Opt. Soc. Am. 66, 1111A.

29. J.R. Fienup (1982). "Phase Retrieval Algorithms: A Comparison," Appl. Opt. 21, 2758-2769.

30. H.V. Deighton, M.S. Scivier and M.A. Fiddy (1985). "Solution of the Two-Dimensional Phase Retrieval Problem," Opt. Lett. 10, 250-251.

31. H.H. Arsenault and S. Lowenthal (1969). "La Restitution de la Phase a de Mesures D'eclairement," C.R. Acad. Sc. Paris 269, 518-521.

32. H.H. Arsenault and K. Chalasinska-Macukow (1983). "The Solution to the Phase Retrieval Problem Using the Sampling Theorem," Opt. Commun. 47, 380-386.

33. K. Chalasinska-Macukow and H.H. Arsenault (1985). "Fast Iterative Solution to Exact Equations for the Two-Dimensional Phase-Retrieval Problem," J. Opt. Soc. Am. A 2. 46-50.

34. R.H.T. Bates, et al. (1982). "Fourier Phase Problems Are Uniquely Solvable in More than One Dimension. I: Underlying Theory," Optik 61, 247-262; (1982) "...II: One-Dimensional Considerations," Optik 62, 131-142; (1982) "... III: Computational Examples for Two Dimension," Optik 62, 219-230.

35. R.W. Gerchberg and W.O. Saxton (1972). "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," Optik 35, 237-246.

36. R.W. Gerchberg (1974). "Super-Resolution through Error Energy Reduction," Optica Acta 21, 709-720.

37. W.O. Saxton (1978). "Computer Techniques for Image Processing in Electron Microscopy" (Academic Press, New York).

38. J.R. Fienup (1979). "Space Object Imaging Through the Turbulent Atmosphere," Opt. Eng. 18, 529-534.

39. J.R. Fienup (1981). "Reconstruction and Synthesis Applications of an Iterative Algorithm," in Transformations in Optical Signal Processing, W.T. Rhodes, J.R. Fienup, and B.E.A. Saleh, eds., Proc. SPIE 373, 147-160.

40. J.R. Fienup (1984). "Phase Retrieval from a Single Intensity Distribution," in Optics in Modern Science and Technology, Conference Digest for ICO-13, August 20-24, Sapporo, Japan, 606-609; J.R. Fienup (1985). "Phase Retrieval Using a Support Constraint," IEEE ASSP Workshop on Multidimensional Digital Signal Processing, Leesburg, VA.

41. J.R. Fienup and G.B. Feldkamp (1980). "Astronomical Imaging by Processing Stellar Speckle Interferometry Data," in Applications of Speckle Phenomena, W.H. Carter, ed., Proc. SPIE 243, 95-102.

42. R.J. Sault (1983). "The Fourier Phase Problem," FLEURS Internal Report No. 83/1, University of Sydney.

43. P. vanToorn A.H. Greenaway and A.M.J. Huiser (1984). "Phaseless Object Reconstruction," Optica Acta 7, 767-774.

44. J.R. Fienup (1984). "Experimental Evidence of the Uniqueness of Phase Retrieval from Intensity Data," in Indirect Imaging, Proceedings of URSI/IAU Symposium, 30 Aug. to 2 Sept. 1983, Sydney, Australia, ed. J.A. Roberts (Cambridge University Press, Cambridge), 99-109.

45. J.R. Fienup and C.C. Wackerman (1984). "Phase Retrieval Stagnation Problems and Solutions," submitted to J. Opt. Soc. Am.; and "Improved Phase Retrieval Algorithms," presented at the October 1984 Annual Meeting of the Optical Society of America, San Diego, CA, Abstract: J. Opt. Soc. Am. 1, 1320A.

46. G.B. Feldkamp and J.R. Fienup (1980). "Noise Properties of Images Reconstructed from Fourier Modulus," in 1980 International Optical Computing Conference, W.T. Rhodes, ed., Proc. SPIE 231, 84-93.

47. A. Labeyrie (1970). "Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images," Astron. and Astrophys. 6, 85-87.

48. J.W. Goodman and J.F. Belsher (1976). "Fundamental Limitations in Linearly Invariant Restoration of Atmospherically Degraded Images," in Imaging Through the Atmosphere, Proc. SPIE 75, 141-154.

49. C.R. Lynds, S.P. Worden and J.W. Harvey (1976). "Digital Image Reconstruction Applied to Alpha Orionis," Astrophys. J. 207, 174-180.

50. R.H.T. Bates and F.M. Cady (1980). "Towards True Imaging by Wideband Speckle Interferometry," Opt. Commun. 32, 365-369.

51. B.R. Hunt, W.R. Fright and R.H.T. Bates (1983). "Analysis of the Shift-And-Add Method for Imaging Through Turbulent Media," J. Opt. Soc. Am. 73, 456.

52. K.T. Knox and B.J. Thompson (1974). "Recovery of Images from Atmospherically Degraded Short-Exposure Photographs," Astrophys. J. Letters 193, L45-L48.

53. M. Rimmer (1974). "Method for Evaluating Lateral Shearing Interferograms," Appl. Opt. 13, 623-629.

54. B.R. Hunt (1979). "Matrix Formulation of the Reconstruction of Phase Values from Phase Differences," J. Opt. Soc. Am. 69, 393-399.

55. A.W. Lohmann, G. Weigelt and B. Wirnitzer (1983). "Speckle Masking in Astronomy: Triple Correlation Theory and Applications," Appl. Opt. 22, 4028-4037.

56. H. Bartelt, A.W. Lohmann and B. Wirnitzer (1984). "Phase and Amplitude Recovery from Bispectra," Appl. Opt. 23, 3121-3129.

57. B.J. Brames and J.C. Dainty (1981). "Method for Determining Object Intensity Distributions in Stellar Speckle Interferometry," J. Opt. Soc. Am. 71, 1542-1545.

58. J.G. Walker (1981). "The Phase Retrieval Problem: A Solution Based on Zero Location by Exponential Apodization," Optica Acta 28, 735-738.

59. J.G. Walker (1982). "Computer Simulation of a Method for Object Reconstruction from Stellar Speckle Interferometry Data," Appl. Opt. 21, 3132-3137.

60. B.L. McGlamery (1970). "Image Restoration Techniques Applied to Astronomical Photography," in V.R. Boscarino, ed., Astronomical Use of Television-Type Image Sensors NASA SP-256(NTIS N71-28522), 167-192.

61. M.S. Scivier, T.J. Hall and M.A. Fiddy (1984). "Phase Unwrapping Using the Complex Zeros of a Bandlimited Function and the Presence of Ambiguities in Two Dimensions," Opt. Acta 31, 619-623.

62. L. Mertz (1984). "Phase Estimation with a Few Photons," Appl. Opt. 23, 1638-1641.

63. W.T. Rhodes and J.W. Goodman (1973). "Interferometric Technique for Recording and Restoring Images Degraded by Unknown Aberrations," J. Opt. Soc. Am. 63, 647-657.

64. T.M. Brown (1978). "Reconstruction of Turbulence Degraded Images Using Nonredundant Aperture Arrays," J. Opt. Soc. Am. 68, 883-889.

65. A.H. Greenaway (1982). "Diffraction-Limited Pictures from Single Turbulence Degraded Images in Astronomy," Opt. Commun. 42, 157-161.

66. R.A. Gonsalves (1982). "Phase Retrieval and Diversity in Adaptive Optics," Opt. Eng. 21, 829-832.

67. L. Mertz (1965). "Transformations in Optics," Wiley, New York.

68. C. Roddier and F. Roddier (1979). In "Image Formation from Coherence Functions in Astronomy," D. Reidel, Amsterdam, pp. 175-185.

69. R.C. Jennison (1958). "A Phase Sensitive Interferometer Technique for the Measurement of the Fourier Transforms of Spatial Brightness Distributions of Small Angular Extent," Mon. Not. R. Astr. Soc. 118, 276-284.

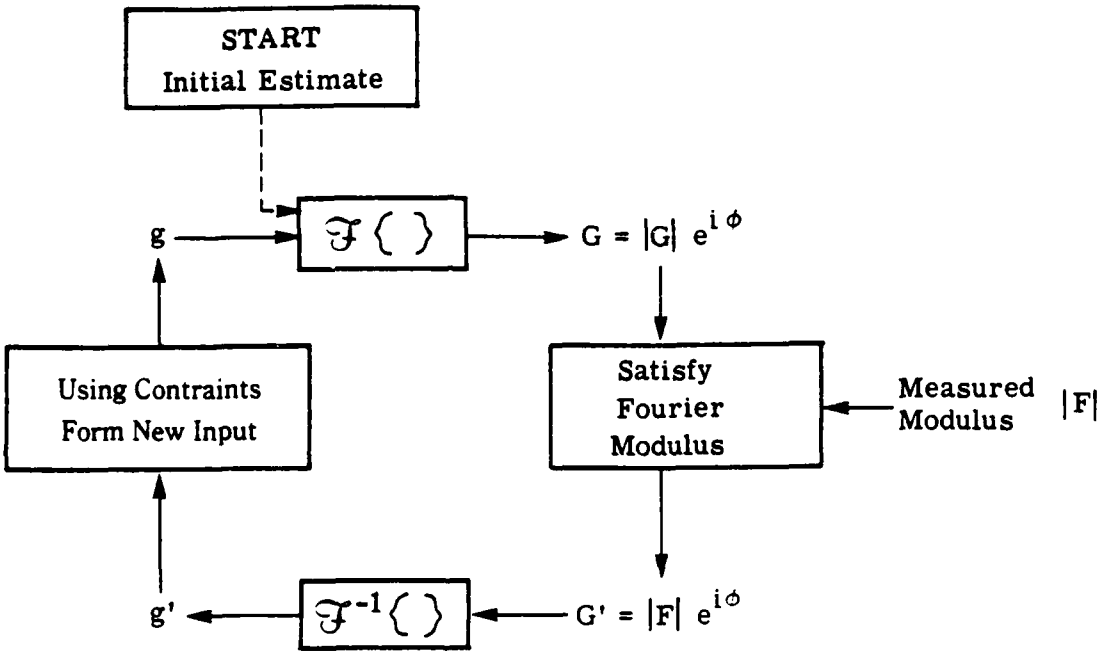


Figure 7.4-1 - Block diagram of the iterative transform algorithm.

Figure 7.4-2 - Example of stagnation with simultaneous twin images. (a) Object f(x,y); (b) twin image f(-x,-y); (c) stagnated output image partially reconstructed from Fourier modulus, having features of both twins.

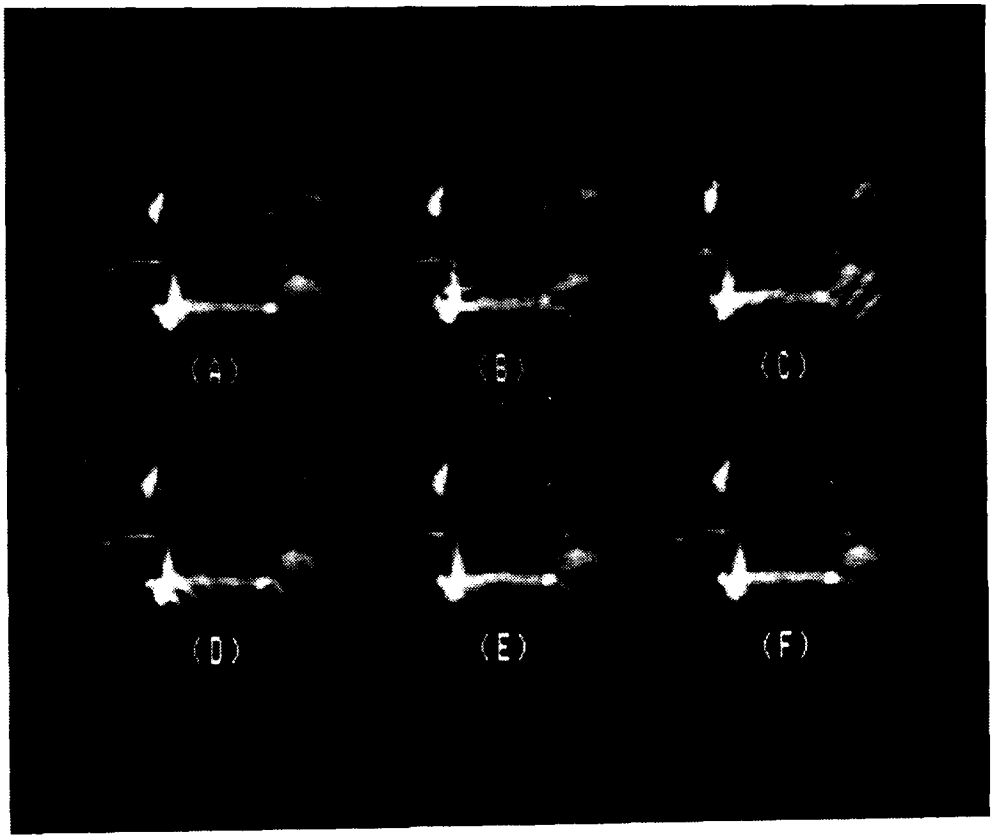


Figure 7.4-3 - Example of stagnation with stripes, and overcoming it by the voting method. (a) object; (b), (c), (d): images partially reconstructed by the iterative transform algorithm each from a different random starting input and each having a different set of stripes superimposed on an otherwise good-quality image; (e) output of the voting method; (f) output image after further iterations.

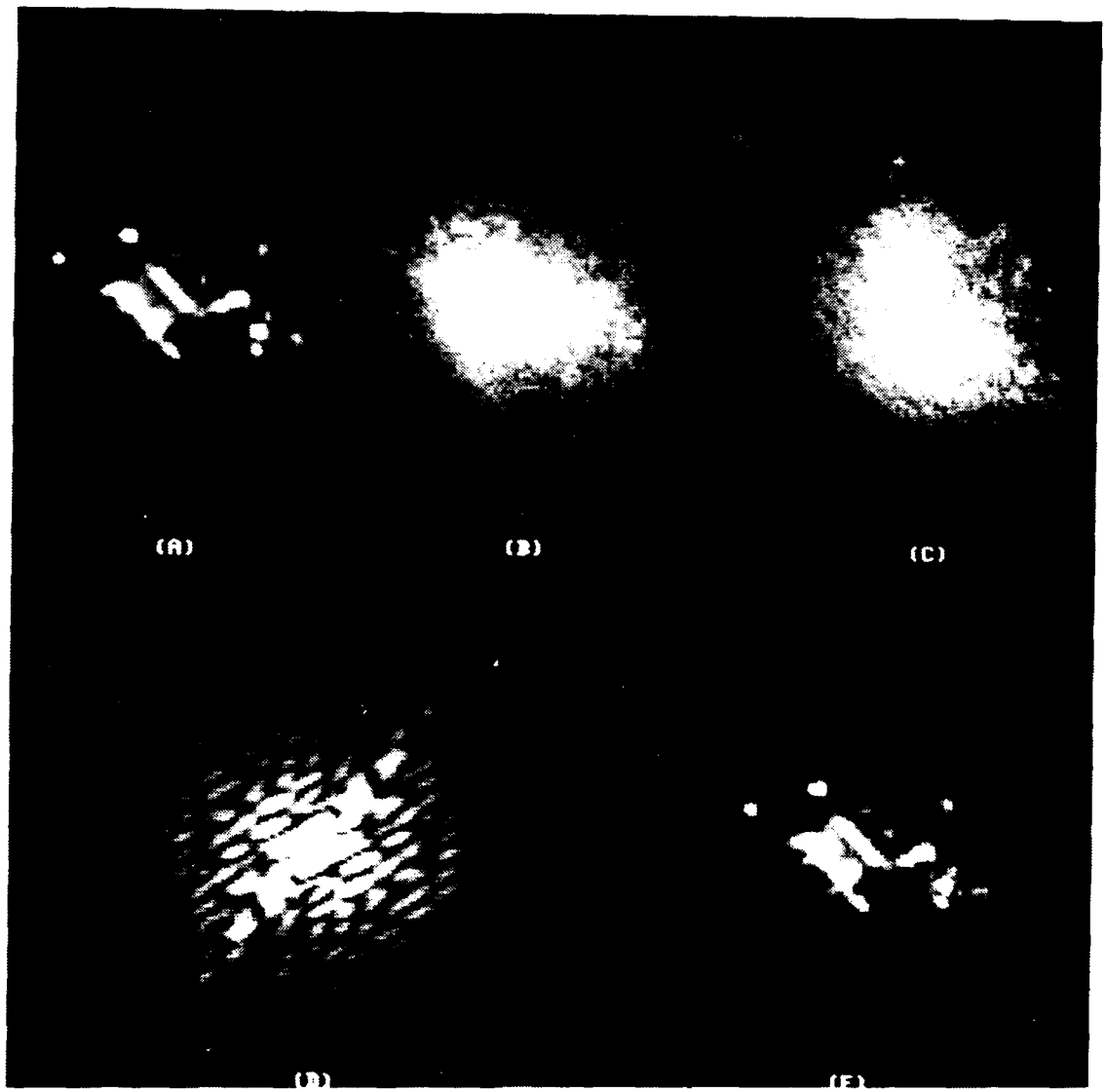


Figure 7.4-4 - Image reconstruction example. (a)Undegraded object; (b), (c) examples of degraded images simulated to include the effects of atmospheric turbulence and photon noise; (d) Fourier modulus estimate computed from the degraded images; (e) image reconstructed using the iterative transform algorithm.

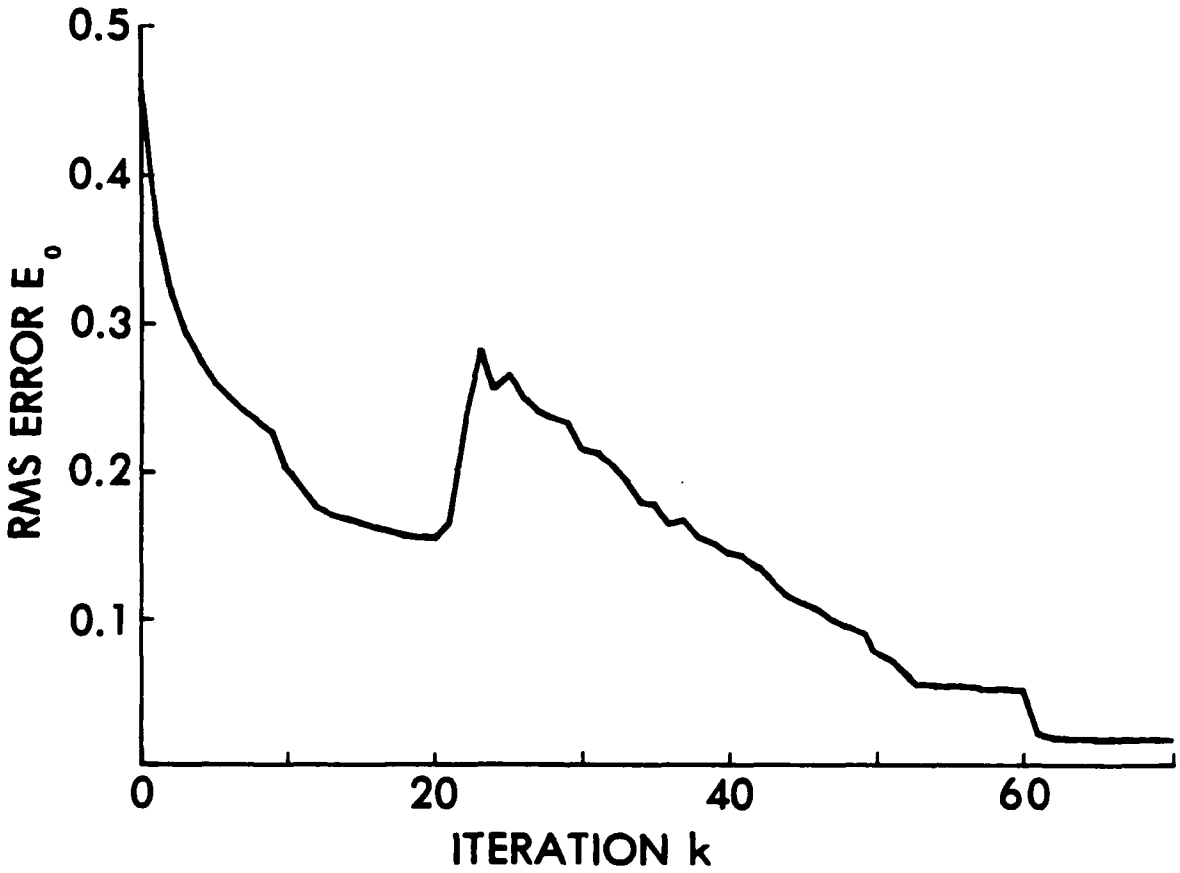


Figure 7.4-5 - Normalized root-mean-squared (NRMS) error metric, E_o, versus iteration number for the example.

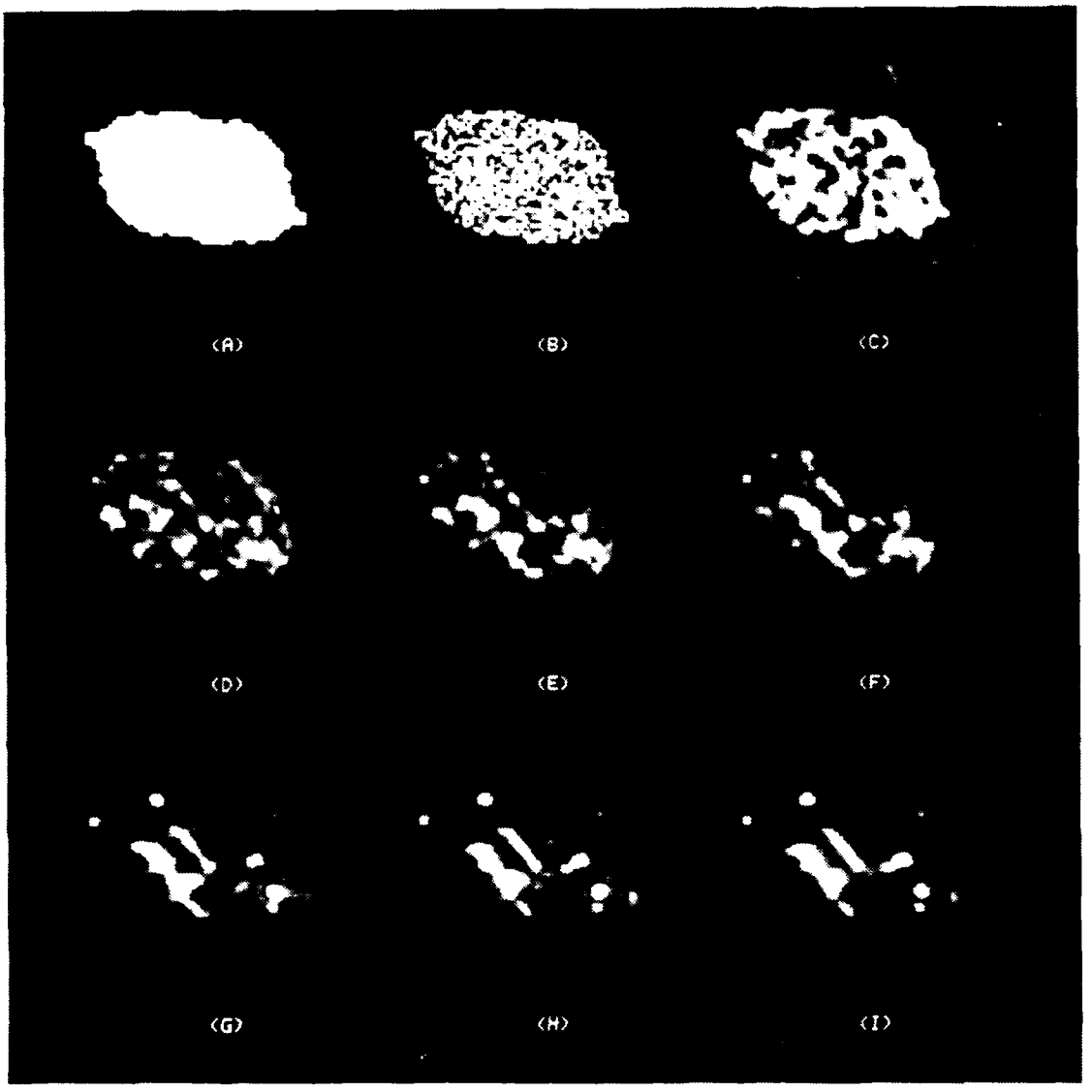


Figure 7.4-6 - Details of the image reconstruction example: (a) mask array defining the support constraint for the first 10 iterations; (b) initial input to the iterative transform algorithm; (c)-(i) output images -- number of iterations: (c) 1, (d) 5, (e) 20, (f) 25, (g) 35, (h) 45, (i) 55.

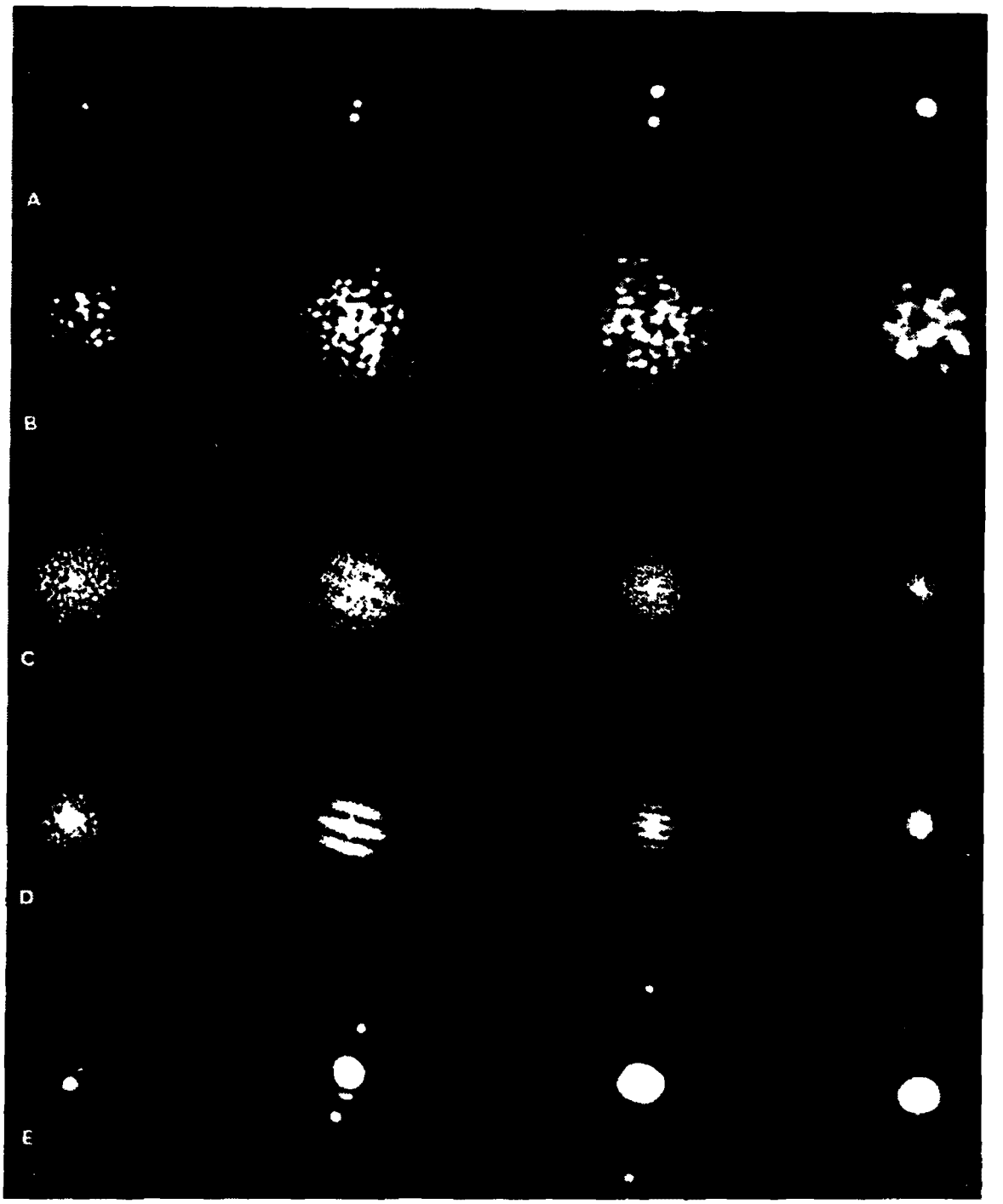


Figure 7.5-1 - Laboratory simulation showing the principles of stellar speckle interferometry. A - objects; B - typical short exposure photographs; C - diffraction patterns of row B; D - sum of 20 diffraction patterns; E - diffraction patterns of row D (not at the same scale as row A). (Courtesy of A. Labeyrie, CERGA).

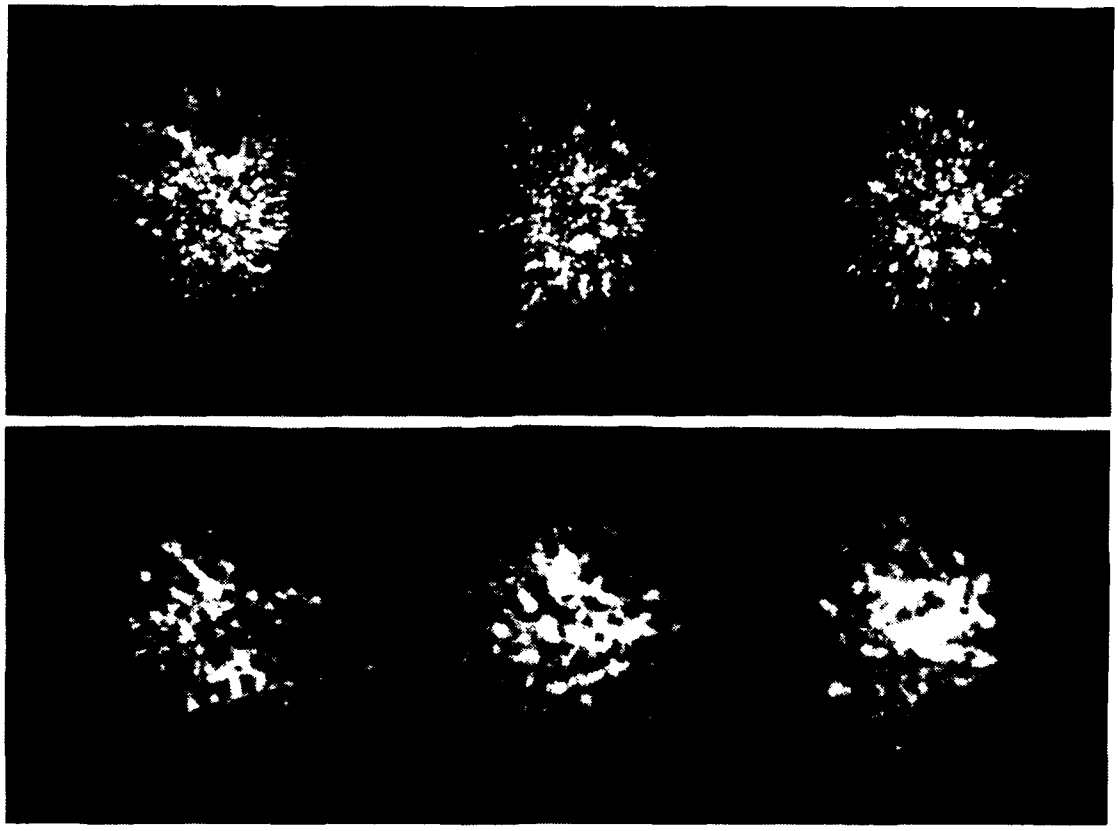


Figure 7.5-2 - Short exposure photographs of an unresolved point source (upper row) and a resolved star, α-Orionis (lower row), taken on a 4 m telescope. The exposure time and filter bandwidth are 10^{-2} s and 10 nm respectively. (Courtesy of B.L. Morgan and R.J. Scaddan, Imperial College, London).

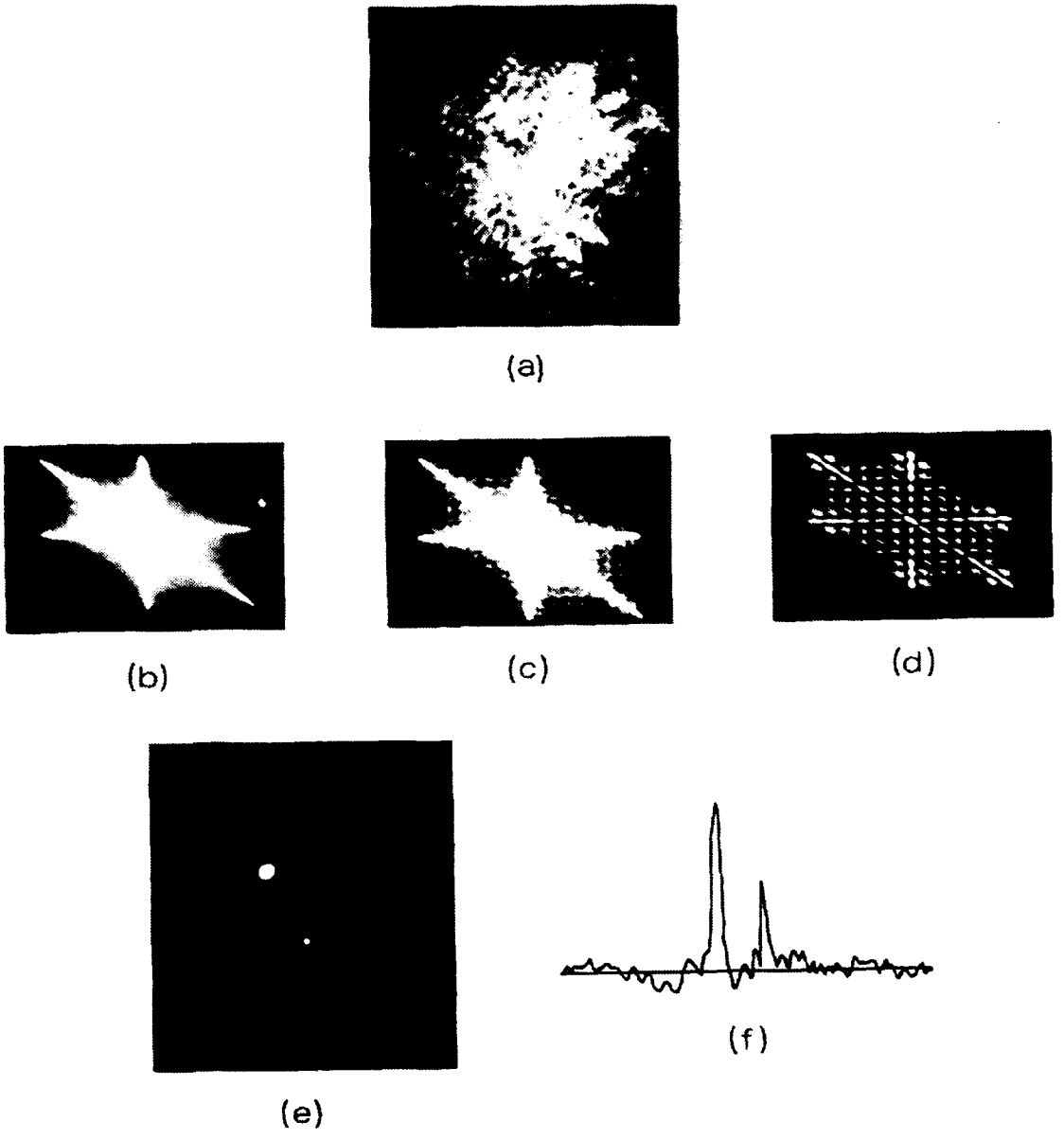


Figure 7.5-3 - Illustration of the triple correlation method in one dimension on the double star ω-Leonis. (a) typical short exposure speckle photograph; (b) bispectrum transfer function; (c) average bispectrum of speckle image calculated from 300 exposures; (d) object bispectrum found from (b) and (c); (e) reconstructed image; (f) scan through reconstructed image. (Courtesy of A.W. Lohmann and G. Weigelt, University of Erlangen).

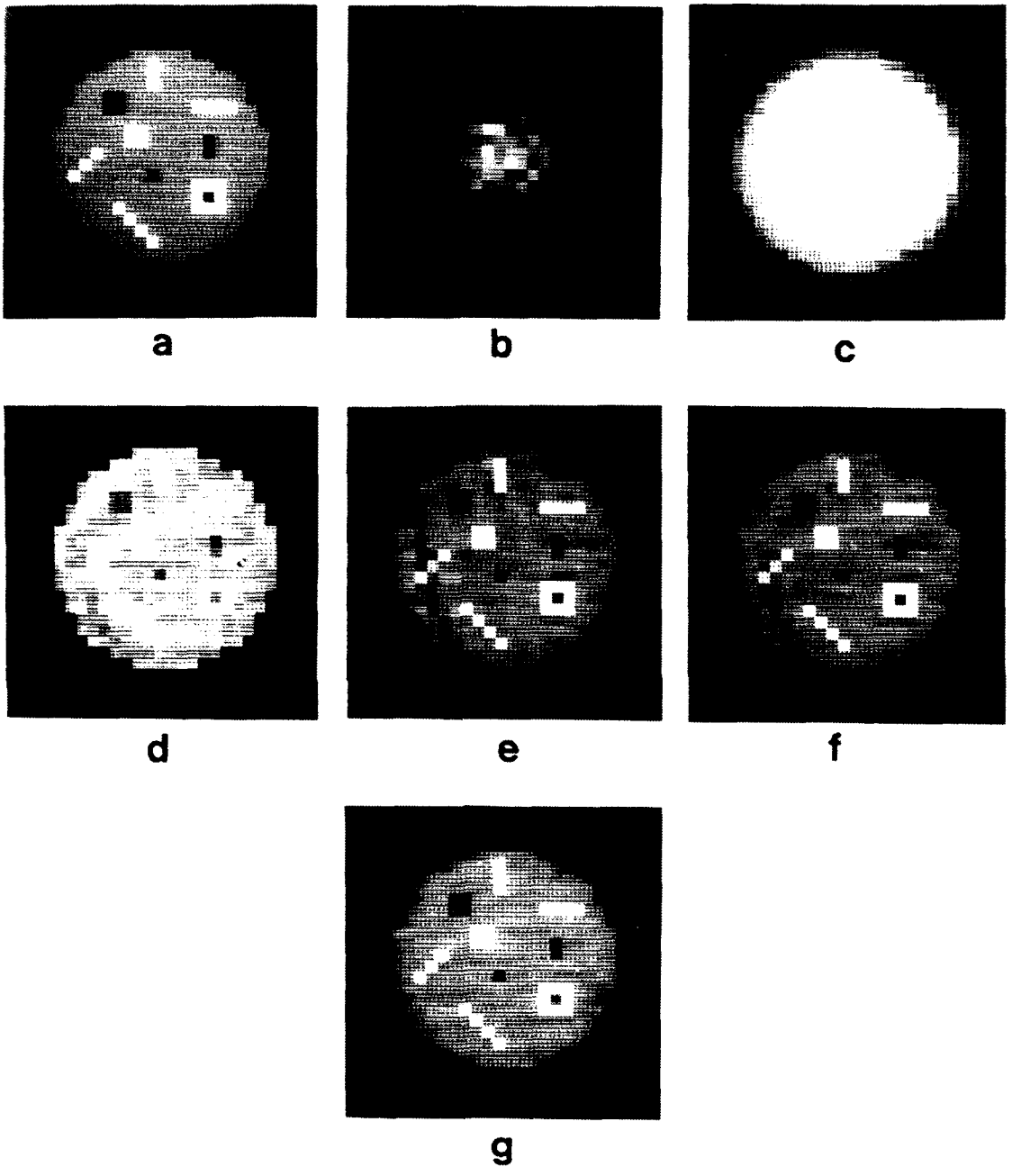


Figure 7.5-4 - Illustration of the exponential filter method. (a) test object; (b) a short exposure point spread function; (c) a short exposure image; (d) to (g) reconstructions using respectively 10, 25, 50 and 100 simulated point and object exposures (Courtesy of J.W. Walker, Kings College, London).

Appendix B

RECONSTRUCTION OF OBJECTS HAVING LATENT REFERENCE POINTS

J.R. Fienup

Reprinted from the Journal of the Optical Society of America 73, 1421-1426 (November 1983).

Reprinted from Journal of the Optical Society of America, Vol. 73, page 1421, November 1983

Reconstruction of objects having latent reference points

J. R. Fienup

Environmental Research Institute of Michigan, P.O. Box 8618, Ann Arbor, Michigan 48107

Received March 18, 1983; revised manuscript received July 14, 1983

A simple recursive algorithm is proposed for reconstructing certain classes of two-dimensional objects from their autocorrelation functions (or equivalently from the modulus of their Fourier transforms—the phase-retrieval problem). The solution is shown to be unique in some cases. The objects contain reference points not satisfying the holography condition but satisfying weaker conditions. Included are objects described by Fiddy *et al.* [Opt. Lett. 8, 96 (1983)] satisfying Eisenstein's thorem.

INTRODUCTION

In a number of disciplines, including astronomy, x-ray crystallography, electron microscopy, and wave-front sensing, one encounters the phase-retrieval problem. One wishes to reconstruct $f(m, n)$, an object function, from $|F(p, q)|$, the modulus of its Fourier transform, where

$$F(p, q) = |F(p, q)|\exp[i\psi(p, q)] = \mathcal{F}[f(m, n)]$$
$$= \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} f(m, n)\exp[-i2\pi(mp/M + nq/N)], \quad (1)$$

where $m, p = 0, 1, \ldots, M-1$ and $n, q = 0, 1, \ldots, N-1$. The discrete transform is employed here since in practice one deals with sampled data in a computer. The problem of reconstructing the object from its Fourier modulus is equivalent to reconstructing the Fourier phase, $\psi(p, q)$, from the Fourier modulus; since once one has the phase as well as the modulus, one can easily compute $f(m, n)$ by the inverse (discrete) Fourier transform. $r_f(m, n)$, the (aperiodic) autocorrelation of $f(m, n)$, is given by[1]

$$r_f(m, n) = \sum_{j=0}^{M-1}\sum_{k=0}^{N-1} f(j, k)f^*(j - m, k - n) \quad (2)$$

$$= \mathcal{F}^{-1}[|F(p, q)|^2], \quad (3)$$

where the asterisk denotes complex conjugate. Note that the autocorrelation is Hermitian: $r_f(-m, -n) = r_f^*(m, n)$. Note also that in order to avoid aliasing during the computation of $|F(p, q)|^2$, it is necessary to have $f(m, n) = 0$ for $M/2 \le m \le M - 1$ and for $N/2 \le n \le N - 1$; this will be assumed throughout this paper. Then there is no difference between the periodic (cyclic) and aperiodic autocorrelation. (For x-ray crystallography this is usually not the case, and the results of this paper do not apply.) Since the autocorrelation function is easily computed from the Fourier modulus by Eq. (3), the phase-retrieval problem is equivalent to reconstructing an object from its autocorrelation function.

Several phase-retrieval algorithms have been proposed, all of them requiring some additional measurements or constraints on the solution. Examples include a reference point at least one object diameter from the object[2] (giving rise to the holography condition[3]), a second intensity measurement in another plane[4,5] (in electron microscopy or wave-front sensing), nonnegativity and limited spatial extent[6-8] (in astronomy), atomic models[9] (in x-ray crystallography), and objects consisting of collections of points having nonredundant spacings.[10]

Here it is pertinent to review the case of holography. Suppose that $f(m, n)$ consists of an object of interest, $g(m, n)$, plus an unresolved (delta-function-like) point, referred to as the reference point, i.e.,

$$f(m, n) = A\delta(m - m_0, n - n_0) + g(m, n), \quad (4)$$

where $\delta(m, n)$ is a two-dimensional (2-D) Kronecker delta function. Then the autocorrelation can be written as the sum of four terms,

$$r_f(m, n) = |A|^2\delta(m, n) + r_g(m, n) + Ag^*(m_0 - m, n_0 - n) + A^*g(m + m_0, n + n_0), \quad (5)$$

the final term of which is the cross-correlation of the reference point with the object of interest and is simply proportional to a translate of the object of interest. If the distance from the reference point to the object of interest exceeds the diameter of the object of interest, then the fourth term in Eq. (5) is nonoverlapping with the other terms, and the object of interest is reconstructed by simple inspection of the autocorrelation. Then the holography condition is satisfied.[2,3] If the amplitude and position of the reference point are unknown (except that the reference point satisfies the holography condition), then the object can be reconstructed only to within a complex factor A^* and to within a translation, and there would be a twofold ambiguity as to whether the object is given by the fourth term or the third term (the conjugate image) of Eq. (5).

In this paper we describe an algorithm for reconstructing certain objects having reference points that do not satisfy the holography condition. For these cases the reference points may be referred to as latent reference points, because they do not immediately yield the object as would a holographic reference point; rather, a degree of development is required before their usefulness emerges.

In Section 2 the question of the uniqueness of the solution is reviewed. In Section 3 the new reconstruction algorithm is described as it is applied to three different classes of objects. Additional comments on the reconstruction algorithm are included in Section 4.

0030-3941/83/111421-06$01.00

2. UNIQUENESS OF THE SOLUTION

When one measures only the Fourier modulus, then the uniqueness of the solution is a central question. One of course always has the twofold (180° rotated or conjugate image) ambiguity since $|\mathcal{F}[f(m,n)]| = |\mathcal{F}[f^*(-m,-n)]|$; and translations of $f(m,n)$ and the multiplication of $f(m,n)$ by a constant phase factor $\exp(i\theta)$ (where θ is a real constant) also have no effect on $|F(p,q)|$. If these are the only ambiguities, then we consider the solution of the phase-retrieval problem to be unique.

Bruck and Sodin[11] considered objects consisting of a rectangular grid of delta functions having various complex amplitudes (or equivalently, a 2-D sequence), which have Fourier transforms that can be expressed as polynomials. These are the types of objects assumed by Eqs. (1) and (2), and we refer to such objects as sampled objects. They showed that, for sampled objects, a lack of uniqueness of the solution to the phase-retrieval problem is equivalent to the factorability of the polynomial, and therefore one-dimensional (1-D) objects of length L have a 2^{L-1}-fold ambiguity.[11] This result corresponds to the analogous theory for 1-D continuous functions.[12] On the other hand, polynomials of two (or more) variables are known to be only rarely factorable (i.e., they are usually irreducible). Consequently, for 2-D sampled objects the solution to the phase-retrieval problem is usually unique. An analogous theory for 2-D continuous functions is not yet available.

Uniqueness Condition Due to Eisenstein's Theorem

Although most 2-D sampled objects are, as discussed above, uniquely related to the modulus of their Fourier transforms, it is of interest to know conditions that ensure uniqueness. Such a condition was recently put forward by Fiddy *et al.*[13] They considered the class of sampled objects whose support is contained in the union of a rectangle and an isolated point (A) below and to the right of the rectangle, as shown in Fig. 1(a). By way of example, the rectangular region in Fig. 1(a)

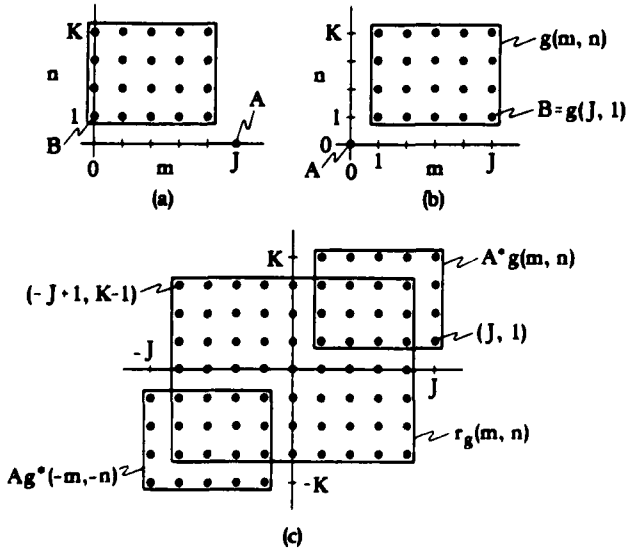


Fig. 1. Fiddy–Brames–Dainty[13] object. (a) FBD object support having two reference points, A and B; (b) object support assumed; (c) autocorrelation support. The object is uniquely reconstructed from its autocorrelation function.

contains five columns and four rows of points. The object must also be nonzero both at point A and at point B in the lower left corner of the rectangle. Points A and B are referred to as the reference points, and they do not satisfy the holography condition. If these conditions are satisfied, then the Fourier transform of the object satisfies Eisenstein's theorem, making it an irreducible 2-D polynomial and guaranteeing that the solution to the phase retrieval problem is unique. They demonstrated the power of these conditions by reconstruction experiments using the input–output iterative Fourier-transform algorithm.[6,7] First, they performed a reconstruction experiment on the Fourier modulus of a particular object that did not have a reference point A. After 250 iterations, a poor reconstruction resulted. But when a new object was formed by adding a reference point A off its corner making it satisfy the conditions, then a good reconstruction was obtained after only 20 iterations.[13] Note that this does not prove that the original object (without the point A) was nonunique: the failure of the iterative reconstruction algorithm may only be an indication of local minima in the error function. In fact, when the reference point A had a small value, a poor reconstruction was obtained in spite of the fact that irreducibility (and uniqueness) was ensured. Only when a large value for A was used did the reconstruction become easier.[13] Apparently the use of a large enough value for A also ensures that there are no local minima.

3. NEW RECONSTRUCTION ALGORITHM

For certain classes of sampled objects having reference points not satisfying the holography condition, we present a new reconstruction algorithm having a fixed number of steps. This new algorithm is related to the Dallas[5] recursive algorithm for phase retrieval from two intensity measurements but requiring only a single intensity measurement (the Fourier modulus) and solving the equations in a certain order such that the problem of a growing tree of solutions[5] is avoided. First the algorithm will be described for the type of object described above, and later for a wider class of objects.

A. Fiddy–Brames–Dainty Objects

For mathematical simplicity, consider a sampled object whose support is contained in the regions shown in Fig. 1(b). Its uniqueness properties are the same as those of the objects considered in Fig. 1(a) since the supports are mirror images of one another. The object can be expressed as in Eq. (4) with $m_0 = n_0 = 0$:

$$f(m,n) = A\delta(m,n) + g(m,n),$$

where $g(m,n)$ is that part of $f(m,n)$ contained in the rectangular region of support, and $A = f(0,0) \neq 0$. In this case, $g(m,n)$ is zero outside $1 \le m \le J$ and $1 \le n \le K$; and it is assumed that $f(J,1) = g(J,1) = B \neq 0$, and $g(m,K) \neq 0$ for at least one value of m. We will refer to objects satisfying these constraints as Fiddy–Brames–Dainty (FBD) objects having FBD regions of support.

The autocorrelation, $r_f(m,n)$, of $f(m,n)$ is given by the four terms of Eq. (5) with $m_0 = n_0 = 0$, the supports of which are contained in the sets of points illustrated in Fig. 1(c). From this figure, it can be clearly seen that the rightmost column and the uppermost row of $r_f(m,n)$ are simply equal to $A^*g(m,n)$:

$$r_f(J, n) = A^* g(J, n), \qquad n = 1, \ldots, K, \tag{6}$$

$$r_f(m, K) = A^* g(m, K), \qquad m = 1, \ldots, J. \tag{7}$$

Therefore, for $m = J$ and for $n = K$, one can reconstruct $g(m, n)$ to within a constant factor A^* by simple inspection of $r_f(m, n)$. In effect, the holography condition is in force for the row and column opposite reference point A, and that row and that column are reconstructed by using reference point A.

The value of A can be obtained as follows: From Eq. (2), it is seen that there is only one nonzero term in the summation for the upper left corner point in the autocorrelation:

$$r_f(-J + 1, K - 1) = g(1, K)g^*(J, 1) = B^* g(1, K). \tag{8}$$

Also, from Eqs. (6) and (7),

$$r_f(J, 1) = A^* g(J, 1) = A^* B, \tag{9}$$

$$r_f(1, K) = A^* g(1, K). \tag{10}$$

Combining Eqs. (8)–(10) yields, assuming that $r_f(-J + 1, K - 1) \neq 0$,

$$|A|^2 = \frac{r_f(J, 1) r_f^*(1, K)}{r_f^*(-J + 1, K - 1)}. \tag{11}$$

Since without loss of generality we can arbitrarily fix the phase of any one point in $f(m, n)$, we set the phase of A equal to zero; A is then given unambiguously by the positive square root of Eq. (11). If $r_f(-J + 1, K - 1) = 0$, then one can obtain a similar expression for $|A|^2$ using the first nonzero point, $r_f(m, K - 1)$, to the right of $r_f(-J + 1, K - 1)$. Since A is known, $g(J, n)$ and $g(m, K)$ can be determined unambiguously from Eqs. (6) and (7). Note that $B = g(J, 1) = r_f(J, 1)/A^*$.

Having the values of the top row and rightmost column of $g(m, n)$, one can then solve for the leftmost column in the second step of the algorithm. From Eq. (2), the point of the autocorrelation just below $r_f(-J + 1, K - 1)$ has only two nonzero terms,

$$r_f(-J + 1, K - 2) = g(1, K)g^*(J, 2) + g(1, K - 1)g^*(J, 1). \tag{12}$$

Solving,

$$g(1, K - 1) = [r_f(-J + 1, K - 2) - g(1, K)g^*(J, 2)]/B^*, \tag{13}$$

where $g(J, 1) = B$. Since all the quantities of the right-hand side of Eq. (13) are known and $B \neq 0$, one can unambiguously compute $g(1, K - 1)$. Similarly, the next lower point in the autocorrelation is given by

$$r_f(-J + 1, K - 3) = g(1, K)g^*(J, 3) + g(1, K - 1)g^*(J, 2) + g(1, K - 2)g^*(J, 1). \tag{14}$$

Since all the quantities in this linear equation are known except for $g(1, K - 2)$, and since $g(J, 1) \neq 0$, one can solve unambiguously for $g(1, K - 2)$. In a similar fashion, one can recursively solve for all the values $g(1, n)$ (the first column on the left) using the values of $r_f(-J + 1, n - 1)$ in this second step of the reconstruction. In a sense the column $m = 1$ was solved using the latent reference point B, which required the solution of column $m = J$ before it could become effective.

Having the first column on the left and the first column on the right of $g(m, n)$, one can then solve for the second column on the right in the third step, using A as the latent reference point. From Eq. (2), the points of the autocorrelation in column $(J - 1)$ are given by

$$r_f(J - 1, n) = g(J - 1, n)A^* + \sum_{k=n+1}^{K} g(J, k)g^*(1, k - n), \tag{15}$$

for $n = 1, \ldots, K - 1$. Since, for any n, $g(J - 1, n)$ is the only unknown in Eq. (15), and since $A \neq 0$, $g(J - 1, n)$ is uniquely determined from Eq. (15). Thus the values of $g(m, n)$ in column $(J - 1)$ are reconstructed using the values in column $(J - 1)$ of the autocorrelation.

The reconstruction algorithm continues in the manner described above. In the fourth step, one can recursively solve for $g(2, n)$ using the latent reference point B and the values of $r_f(-J + 2, n - 1)$, $n = K - 1, K - 2, \ldots, 2, 1$. In the fifth step, one can solve for $g(J - 2, n)$ using the latent reference point A and the values of $r_f(J - 2, n)$, $n = 1, \ldots, K - 1$. One continues the procedure until all the columns of $g(m, n)$ are reconstructed, giving a complete and unambiguous reconstruction of $g(m, n)$, and therefore of $f(m, n)$.

If $g(1, K) \neq 0$, then one can alternatively use that point as B and perform the reconstruction as described above, but reversing the roles of the rows and columns.

It was recently noted that Eisenstein's theorem allows for the rectangular region of support (see Fig. 1) to be extended over (in the same column as) point A. However, in that case, there is no simple recursive algorithm for reconstructing the object.

B. Support Uniqueness for Fiddy-Brames-Dainty Objects

In the reconstruction method described above, it was implicitly assumed that the support of the object function was known. However, as will be shown by what follows, such an assumption is not necessary, since an FBD object can be shown to be an FBD object from its autocorrelation. In order to use theorems[10] relating to reconstructing the support of an object from the support of its autocorrelation function, during the discussion of the object and autocorrelation supports we assume that the object function is real and nonnegative. (It might happen that what follows may, with appropriate modifications, also be true for complex-valued objects; but this would require further development.)

Given only the support of the autocorrelation, one can usually only put an upper bound on the support of the object.[10] Such upper bounds, sets that can contain translates of the supports of all possible solutions, we refer to as locator sets. One such locator set is the intersection of the autocorrelation support with a translate of itself, where the translate is such that the center of the second autocorrelation support is within the first autocorrelation support.[10] Assuming that $r_f(-J + 1, K - 1) \neq 0$, and translating the one autocorrelation support so that it is centered at $(-J + 1, K - 1)$, one arrives at the locator set shown in Fig. 2 for the case of the FBD object support shown in Fig. 1(b). In addition, since the autocorrelation is $2J + 1$ pixels wide and $2K + 1$ pixels high, the object must be $J + 1$ pixels wide and $K + 1$ pixels high. Since the object support must be contained within the locator set shown in Fig. 2, which is $J + 2$ pixels wide and $K + 2$ pixels high, the object support must include either the lower left point or the upper right point but not both. Keeping either one of these

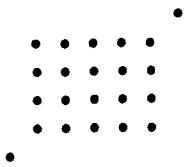


Fig. 2. Locator set containing all possible solutions, used to show that the support solution is unique.

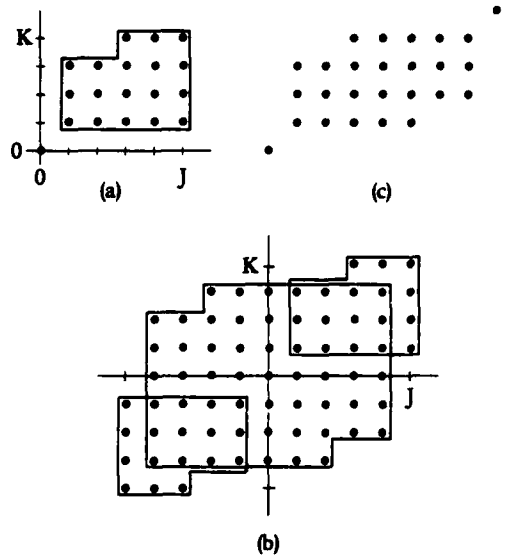


Fig. 3. Alternative case. (a) Object support; (b) autocorrelation support; (c) locator set.

two points and discarding the other, one is left with the support of the object (or the 180° rotated version—the twofold ambiguity). Suppose, on the other hand, that $r_f(-J+1, K-1) = 0$. For example, suppose that the object support is that shown in Fig. 3(a). Then the autocorrelation support is that shown in Fig. 3(b). A locator set, formed by taking the intersection of this autocorrelation support with one translated to be centered at the first nonzero point in row $(K-1)$, is shown in Fig. 3(c). As in the case of Figs. 1 and 2, since the autocorrelation is $2K+1$ pixels high, the object must be $K+1$ pixels high, and therefore either the lower right or the upper left point (but not both) in Fig. 3(c) must be within the object support. Suppose we take the lower left point as being within the object (choosing the upper right point will result in the 180° rotated solution). Then, since the autocorrelation is $2J+1$ pixels wide and therefore the object must be $J+1$ pixels wide, the object must be contained within the first $J+1$ columns on the left of Fig. 3(c), which is just the support of the object as shown in Fig. 3(a). From these examples it can be seen that, in general, if the object is an FBD object, then its support can be reconstructed from the autocorrelation function, from which it is also evident that the object has an FBD support.

C. Triangular Objects

Other types of objects, in addition to FBD objects, can be reconstructed by the recursive method. In this and the next section the reconstruction of two other classes of objects are shown. Consider, for example, objects whose support is contained in the triangular shape shown in Fig. 4(a). Assuming that the object's support is known *a priori*, it has been shown that for this particular object shape the boundaries can be reconstructed in a simple way,[14] assuming $A, B, C \neq 0$. Since the vector spacings between points A and B, B and C, and C and A are each unique, from the corner points in the autocorrelation, as shown in Fig. 4(b), we have

$$r(0, K) = f(0, K)f^*(0, 0) = CA^*, \tag{16a}$$

$$r(J, -K) = f(J, 0)f^*(0, K) = BC^*, \tag{16b}$$

$$r(J, 0) = f(J, 0)f^*(0, 0) = BA^*. \tag{16c}$$

Combining these gives

$$|A|^2 = \frac{r^*(0, K)r(J, 0)}{r(J, -K)}. \tag{17}$$

Without loss of generality the phase of A can be chosen to be zero, and then A is given by the positive square root of Eq. (17). Then we can also compute

$$B = r(J, 0)/A^*, \tag{18a}$$

$$C = r(0, K)/A^*. \tag{18b}$$

Then the values of the leftmost column of the object are given by

$$f(0, n) = r(-J, n)/B^*, \tag{19}$$

the values of the bottom row are given by

$$f(m, 0) = r(m, -K)/C^*, \tag{20}$$

and the values of the diagonal are given by

$$f(m, K-m) = r(m, K-m)/A^*. \tag{21}$$

From this point one could determine the remainder of the object by solving systems of equations,[5,14] but an easier way

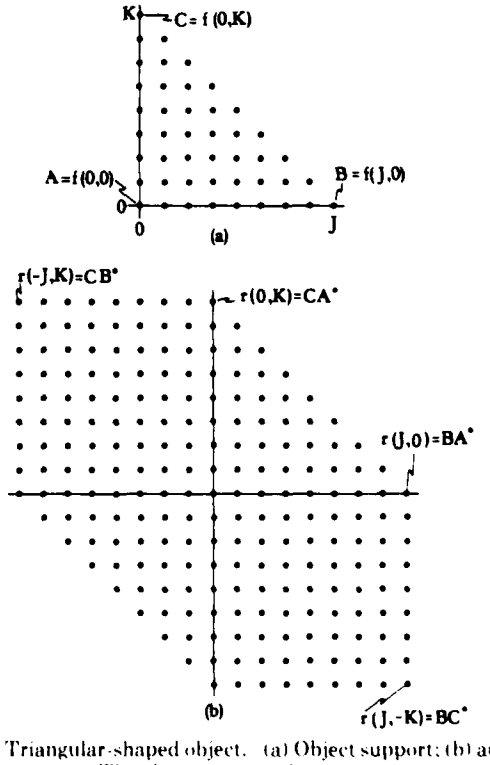


Fig. 4. Triangular-shaped object. (a) Object support; (b) autocorrelation support. The object is uniquely (among triangular-shaped solutions) reconstructed from its autocorrelation function.

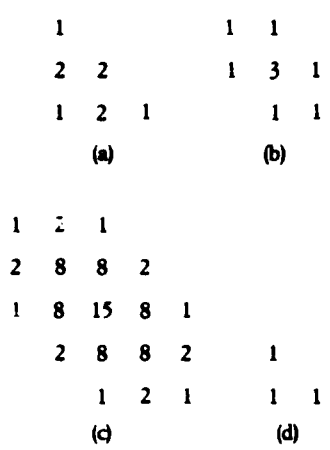


Fig. 5. Specific triangular-shaped object. (a) The object; (b) a second nontriangular-shaped solution; (c) the common autocorrelation function; (d) the function used to synthesize objects shown in (a) and (b).

is possible if one cleverly chooses the order in which the equations are solved. In particular, only one linear equation with one unknown at a time need be solved, and the solution at each step is unique, if one solves in the following order. In a similar manner as was done for the FBD objects, solve for the points in column $m = 1$ using B as a latent reference point, and solve for the points in row $n = 1$ using C as a latent reference point. Next solve for the points in column $m = 2$ using B as a latent reference point, and solve for the points in row $n = 2$ using C as a latent reference point. This procedure is continued until all of $f(m, n)$ is reconstructed. Other orderings for the recursive solution of the equations are also possible.

The solution given above for the triangular-shaped object is unique among objects having that support but may not be unique among all objects. Momentarily restricting $f(m, n)$ to the case of nonnegative objects, one can use the autocorrelation support tri-intersection reconstruction for convex sets[10] to show that there exists a family of object supports that have autocorrelation supports equal to the one shown in Fig. 4(b). One member of that family is the original object support shown in Fig. 4(a). Another member is an object support resembling the autocorrelation support shown in Fig. 4(b) but only half its size. For these latter members there is no simple recursive reconstruction algorithm as there is for the triangular-shaped object.

Further insights can be obtained by analyzing a simple case. A case for which there are exactly two different solutions (not counting 180°-rotated versions) can be obtained by starting with nonsymmetric functions $h_1(x, y)$ and $h_2(x, y)$ whose Fourier transforms are nonfactorable and generating a first object, which is $h_1(x, y)$ convolved with $h_2(x, y)$, and a second object, which is $h_1(x, y)$ convolved with $h_2(-x, -y)$ (i.e., the cross correlation).[15] Two such objects, their common autocorrelation function, and the $h_1(x, y) = h_2(x, y)$ used to generate them are shown in Figs. 5(a) through 5(d), respectively. In this case one obtains the "unique" solution shown in Fig. 5(a) if triangular support is assumed, and the "unique" solution shown in Fig. 5(b) if the only other possible support is assumed.

Since relatively few 2-D objects have factorable Fourier transforms, the ambiguous example shown in Fig. 5 is unusual. If one started with a random object having the same support as the object in Fig. 5(b), and if one incorrectly assumed that the object had the same triangular support as the object in Fig. 5(a), then one would obtain what at first glance would appear to be a triangular-shaped solution. In the process of calculating the solution one would use only the points on the perimeter of the autocorrelation function, with which the "solution" would be consistent. However, on further inspection one would usually find that the triangular-shaped solution is inconsistent with the interior points of the autocorrelation function. Only in the unlikely event that the original object's Fourier transform is factorable would the triangular-shaped solution be completely consistent with the autocorrelation function. Therefore if the given autocorrelation function admits to a possible solution by the recursive method, then one should reconstruct the solution with the assumed support, then compute its autocorrelation function and compare it with the given autocorrelation function to determine whether the assumed support is valid.

D. Another Case

For a final example, consider objects contained within the support shown in Fig. 6(a). Comparing it with Fig. 1(b), it would be a FBD object if it were not for the fact that $B = 0$. Assuming that the support of the object is known, it can be reconstructed by the following recursive steps if points A and $B' \neq 0$ and if either point C' or $C'' \neq 0$. First $f(J, 2), \ldots, f(J, K)$ and $f(2, K), \ldots, f(J-1, K)$ are solved using A as the reference point. A can be determined from an equation similar to Eqs. (11) and (17). Next $C' = f(1, K-1)$, then $f(1, K-2)$, ..., then $f(1, 2)$ are solved using B' as the latent reference point. Next $f(J-1, 1)$ is solved using C' or C'' as the latent reference point. Next $f(1, 1)$ is solved using B' as the latent reference point. Next $f(J-1, 2), \ldots, f(J-1, K-1)$ are solved using A as the latent reference point. Then the pattern repeats: solve for $f(2, K-1), \ldots, f(2, 2)$ recursively using B', then solve for $f(J-2, 1)$ using C' or C'', then solve for $f(2, 1)$

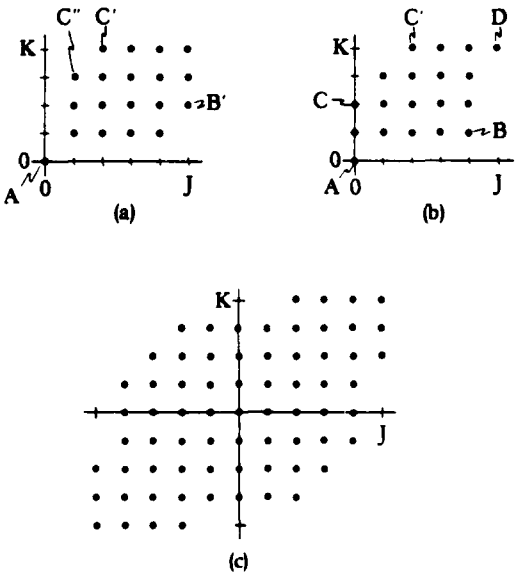
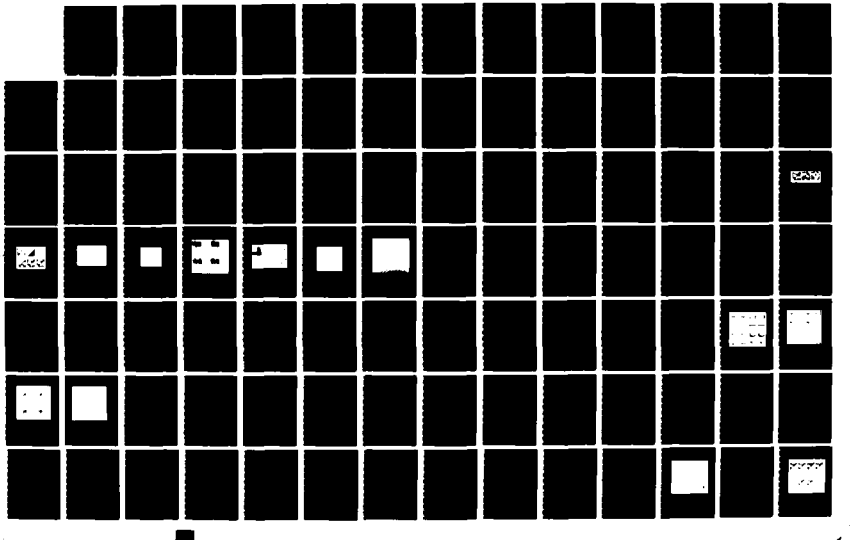
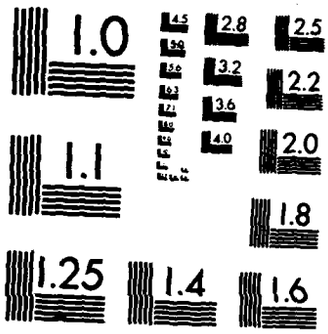


Fig. 6. Another case related to FBD objects. (a) Object support; (b) alternative support reconstruction; (c) autocorrelation support. The object is reconstructed from its autocorrelation function, with two solutions.

AD-A174 100 DIFFRACTION-LIMITED IMAGING OF SPACE OBJECTS III(U) 2/3
ENVIRONMENTAL RESEARCH INST OF MICHIGAN ANN ARBOR
J R FIENUP ET AL OCT 86 ERIM-161900-20-F
UNCLASSIFIED AFOSR-TR-86-2109 F49620-82-K-0018 F/G 14/5 NL





using B', then solve for $f(J-2,2),\ldots,f(J-2,K-1)$ using A, etc., until all the columns are solved.

The solution for this object is unique among objects having support contained within the support shown in Fig. 6(a). However, another support may also be possible. In a manner similar to that used in connection with Figs. 1–3, the possible support solutions can be narrowed down to those of Fig. 6(a) and Fig. 6(b), given the autocorrelation support shown in Fig. 6(c). For the support shown in Fig. 6(b) one can reconstruct the object unambiguously by solving a proper sequence of equations using latent reference points A, B, C, C', and D. Therefore, given the autocorrelation function whose support is shown in Fig. 6(c), at most two (and more probably only one) solutions are possible, and each can be reconstructed using a simple recursive algorithm depending on the support shown in either Fig. 6(a) or 6(b).

4. CONCLUSIONS

A simple recursive algorithm has been devised for reconstructing an object from its autocorrelation function (or its Fourier modulus). It works for several types of sampled objects having latent reference points, including those satisfying the conditions described by FBD. The manner in which the algorithm results in a unique solution constitutes a proof of uniqueness for FBD objects (but not necessarily for all objects whose Fourier transforms satisfy Eisenstein's theorem). One of the principal lessons learned here is that the detailed shape of the boundary of an object plays a crucial role in determining the uniqueness of the solution to the phase-retrieval problem.

One might be able to use this method for continuous objects (as opposed to inherently sampled objects) if a dense enough sampling of the autocorrelation is available.[5]

Since the algorithm involves repeatedly taking differences and dividing by the values of the latent reference points, it may be sensitive to noise and may require latent reference points having large values for an accurate reconstruction. (This may be related to the fact that a large value of A was required for a successful reconstruction using the iterative Fourier-transform algorithm.[13]) Not all the (nonsymmetric) points in the autocorrelation are used by this algorithm; improved accuracy should be expected if the reconstruction algorithm were modified to use also those additional points. Those additional points may also be used to distinguish whether assumptions about the support of the object (when more than one support solution is possible) are valid. For the best results one should finish the reconstruction by using the output of this reconstruction method as the initial input to the iterative Fourier-transform algorithm,[6,7] which finds a solution that is most consistent with both the measured data and the *a priori* constraints.

The reconstruction algorithm proposed here is applicable to only a relatively small number of types of objects. However, the approach of carefully selecting the order in which the equations are solved should be helpful in the more general use of Dallas's method by limiting the growth of the tree of solutions.[5]

ACKNOWLEDGMENT

Helpful discussions with T. R. Crimmins are gratefully acknowledged. This research was supported by the U.S. Air Force Office of Scientific Research under contract F49620-82-K-0018.

REFERENCES

1. R. N. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed. (McGraw-Hill, New York, 1978).
2. C. Y. C. Liu and A. W. Lohmann, "High resolution image formation through the turbulent atmosphere," Opt. Commun. **8**, 372–377 (1973).
3. J. W. Goodman, "Analogy between holography and interferometric image formation," J. Opt. Soc. Am. **60**, 506–509 (1970).
4. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik **35**, 237–246 (1972); W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy* (Academic, New York, 1978); R. H. Boucher, "Convergence of algorithms for phase retrieval from two intensity distributions," Proc. Soc. Photo-Opt. Instrum. Eng. **231**, 130–141 (1980).
5. W. J. Dallas, "Digital computation of image complex amplitude from image- and diffraction-intensity: an alternative to holography," Optik **44**, 45–59 (1975).
6. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982).
7. J. R. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," Opt. Lett. **3**, 27–29 (1978); J. R. Fienup, "Space object imaging through the turbulent atmosphere," Opt. Eng. **18**, 529–534 (1979).
8. P. J. Napier and R. H. T. Bates, "Inferring phase information from modulus information in two-dimensional aperture synthesis," Astron. Astrophys. Suppl. **15**, 427–430 (1974).
9. G. H. Stout and L. H. Jensen, *X-Ray Structure Determination* (Macmillan, London, 1968).
10. J. R. Fienup, T. R. Crimmins, and W. Holsztynski, "Reconstruction of the support of an object from the support of its autocorrelation," J. Opt. Soc. Am. **72**, 610–624 (1982).
11. Yu. M. Bruck and L. G. Sodin, "On the ambiguity of the image reconstruction problem," Opt. Commun. **30**, 304–308 (1979).
12. A. Walther, "The question of phase retrieval in optics," Opt. Acta **10**, 41–49 (1963); E. M. Hofstetter, "Construction of time-limited functions with specified autocorrelation functions," IEEE Trans. Inf. Theory **IT-10**, 119–126 (1964).
13. M. A. Fiddy, B. J. Brames and J. C. Dainty, "Enforcing irreducibility for phase retrieval in two dimensions," Opt. Lett. **8**, 96–98 (1983).
14. M. H. Hayes and T. F. Quatieri, "The importance of boundary conditions in the phase retrieval problem," IEEE Trans. Acoust. Speech Signal Process. **ASSP-82**, 1545 (1982).
15. J. R. Fienup, "Image reconstruction for stellar interferometry," in *Current Trends in Optics*, F. T. Arecchi and F. R. Aussenegg, eds. (Taylor and Francis, London, 1981), pp. 95–102.

Appendix C

PHASE RETRIEVAL STAGNATION PROBLEMS AND SOLUTIONS

J.R. Fienup and C.C. Wackerman

To appear in the Journal of the Optical Society of America A (November 1986).

Appendix C

PHASE RETRIEVAL STAGNATION PROBLEMS AND SOLUTIONS

by

J.R. Fienup and C.C. Wackerman
Environmental Research Institute of Michigan
P.O. Box 8618
Ann Arbor, MI 48107

ABSTRACT

The iterative Fourier transform algorithm has been demonstrated to be a practical method for reconstructing an object from the modulus of its Fourier transform (i.e., solving the problem of recovering phase from a single intensity measurement). In some circumstances the algorithm may stagnate. New methods are described that allow the algorithm to overcome three different modes of stagnation: those characterized by (1) twin images, (2) stripes, and (3) truncation of the image by the support constraint. Curious properties of Fourier transforms of images are also described: the zero-reversal for the striped images and the relationship between the zero lines of the real and imaginary parts of the Fourier transform. A detailed description of the reconstruction method is given to aid those employing the iterative transform algorithm.

To appear in the Journal of the Optical Society of America A (November 1986).

1. INTRODUCTION

In a number of different disciplines including astronomy, wavefront sensing, x-ray crystallography and holography, one encounters the phase retrieval problem: given the modulus $|F(u, v)|$ of the Fourier transform

$$F(u, v) = |F(u, v)| \exp [i\psi(u, v)] = \mathcal{F}[f(x, y)]$$
$$= \iint_{-\infty}^{\infty} f(x, y) \exp [-i2\pi(ux + vy)] \, dx \, dy \qquad (1)$$

of an object $f(x, y)$, reconstruct the object $f(x, y)$ or, equivalently, reconstruct the Fourier phase $\psi(u, v)$. (Here and throughout this paper functions represented by upper case letters are the Fourier transforms of the functions represented by the corresponding lower case letters.) Since the autocorrelation of the object is given by $\mathcal{F}^{-1}[|F(u, v)|^2]$, this is equivalent to reconstructing an object from its autocorrelation. Many solutions to this problem have been proposed [1-12]. The method of solution we feel is most practical from the point-of-view of minimum computational complexity, minimum sensitivity to noise and applicability under the most general assumptions is the iterative Fourier transform algorithm [1-3].

The iterative transform algorithm, a descendant of the Gerchberg-Saxton algorithm [13-15], bounces back and forth between the object domain, where a priori knowledge about the object such as nonnegativity or its support is applied (the support is the set of points over which the object is nonzero), and the Fourier domain, where the measured Fourier modulus data is applied. The algorithm is reviewed in Section 2. Although the algorithm works well for many cases of interest, there is no guarantee that it will converge to a solution.

For certain types of objects the iterative algorithm can stagnate on images which are not fully reconstructed. Stagnation of the

algorithm means that the output image changes very little after many further iterations while not at a solution. A solution is any Fourier transform pair that satisfies the measured data and constraints in both domains with an error metric (defined later) no greater than the expected root-mean-squared error of the measured data. This paper will describe three such conditions of stagnation and algorithms that we have developed to jump each of these stagnation hurdles, allowing the algorithm to move on toward a solution. The three modes of stagnation are those characterized by (1) simultaneous twin images, (2) stripes superimposed on the image, and (3) unintentional truncation by the support constraint.

The first stagnation problem results from the fact that an object $f(x, y)$ and its twin $f^*(-x, -y)$ (the complex conjugated object rotated by 180 degrees) both have the same Fourier modulus and, for cases in which the support of the object is symmetric with respect to this rotation, have the same support. If the iterative algorithm starts from an initial guess of random numbers there is an equal probability that it will reconstruct either of these two objects. The problem arises when, during the initial stages of reconstruction, features of both objects are reconstructed simultaneously. If this situation continues and features of both objects become equally strong (in the sense that applying the constraints does not favor one over the other), the iterative algorithm may stagnate. Not able to suppress one twin image and converge to the other, the algorithm tries to reconstruct both together and goes nowhere.

Since the Fourier transform of the linear combination $t\, f(x, y) + (1 - t)f^*(-x, -y)$ does not have modulus $|F(u, v)|$ except for $t = 1$ or 0 [assuming $f^*(-x, -y) \neq f(x, y)$], an image outputted by the algorithm in this condition is not a solution consistent with the data. This is recognized by the algorithm from the fact that the error metrics (defined in Section 2) are nonzero (or in the presence of noise, greater than the expected RMS error of the data)

in this circumstance. This mode of stagnation often occurs if the support of the object is centro-symmetric. An example of twin-image stagnation is displayed in Figure 2(b) of Reference 16. A method for overcoming this stagnation problem is described in Section 3.

The second stagnation problem is characterized by an output image that looks much like the true object but with a pattern of stripes superimposed. The pattern of stripes is approximately sinusoidal in one direction and constant in the orthogonal direction. The stripes are usually of low contrast and therefore are not objectionable, but they occasionally are of sufficiently high contrast to be disturbing. They are stronger over the support of the object and weaker outside the support. This problem frequently occurs to varying degrees. Examples of its occurence are Figures 3(f), 3(i), 4(a) and 4(b) of Reference 1, Figures 9(a)-9(d) of Reference 17, and to a lesser extent Figure 2(b) of Reference 16. The error metric is nonzero when they are present since the stripes extend (although with lower contrast) outside the known support of the object, and so the striped images are not a solution and do not represent a uniqueness problem. Methods for overcoming this mode of stagnation are given in Section 4. During the study of the stripes phenomenon, some interesting properties of the Fourier transforms were discovered, they too are described in Section 4.

The third stagnation problem arises when the support constraint is used in a manner that is inconsistent with the partially reconstructed image outputted by the algorithm. If the partially reconstructed image is in a position that is translated relative to the position of the mask array defining the support constraint, then the object domain step of the algorithm will inadvertently tend to truncate (cut off spatially) part of the image. This usually results in stagnation of the algorithm. A previously reported method of reducing the likelihood of encountering this problem [3] is described in Section 6. A new method for overcoming this stagnation problem is introduced in Section 5.

As an aid to the practical implementation of the iterative transform algorithm, Section 6 discusses a number of helpful hints that make it converge more reliably. This description should help those employing the iterative transform algorithm to achieve greater success and to convert some of the "black art" of the iterative approach into a more automatic algorithm. Section 7 contains a summary and conclusions.

2. REVIEW OF THE ITERATIVE TRANSFORM ALGORITHM

When working with sampled data on a digital computer, one employs the discrete Fourier transform (DFT)

$$F(u) = \sum_{x=0}^{N-1} f(x) \exp(-i2\pi u\, x/N) \tag{2}$$

and its inverse

$$f(x) = N^{-2} \sum_{u=0}^{N-1} F(u) \exp(i2\pi u\, x/N) \tag{3}$$

which can be computed using the Fast Fourier transform (FFT) method. Here we employ u and x as two-dimensional vectors, $x = (x_1, x_2)$, $u = (u_1, u_2)$, where u_1, u_2, x_1, and $x_2 = 0, 1, 2, \ldots, N - 1$ (square arrays are assumed for simplicity). In order to avoid aliasing in the computation of $|F(u)|^2$, we restrict $f(x)$ to be zero for $x_1 \geq N/2$ and for $x_2 \geq N/2$. Therefore we are considering only problems for which the object has finite support. For problems in astronomy, $f(x)$ is a real, nonnegative function; but for other problems $f(x)$ may be complex-valued. This paper assumes the case of real, nonnegative objects (particularly in the discussion of stripes), although much of the discussion can also be applied to the more general case of complex-valued images.

The simplest version of the iterative transform algorithm follows the philosophy of the Gerchberg-Saxton algorithm, and it is known as the error-reduction algorithm [1]. It can be viewed in a number of different ways: in terms of the method of successive approximations [18], as a form of steepest-descent gradient search [3], or as a projection onto sets in a Hilbert space (the Fourier modulus constraint being onto a nonconvex set, however, so convergence is not assured) [19].

For the most general problem, the error-reduction algorithm consists of the following four steps (for the k^{th} iteration): (1) Fourier transform $g_k(x)$, an estimate of $f(x)$, yielding $G_k(u)$; (2) make the minimum changes in $G_k(u)$ which allow it to satisfy the Fourier-domain constraints to form $G'_k(u)$, an estimate of $F(u)$; (3) inverse Fourier transform $G'_k(u)$, yielding $g'_k(x)$, the corresponding image; and (4) make the minimum changes in $g'_k(x)$ that allow it to satisfy the object-domain constraints to form $g_{k+1}(x)$, a new estimate of the object. For phase retrieval from a single intensity measurement, in which the Fourier modulus $|F(u)|$ is the square root of the intensity, these four steps are

$$G_k(u) = |G_k(u)| \exp [i\phi_k(u)] = \mathcal{F}[g_k(x)], \tag{4}$$

$$G'_k(u) = |F(u)| \exp [i\phi_k(u)], \tag{5}$$

$$g'_k(x) = \mathcal{F}^{-1}[G'_k(u)], \tag{6}$$

$$g_{k+1}(x) = \begin{cases} g'_k(x), & x \notin \gamma \\ 0, & x \in \gamma, \end{cases} \tag{7}$$

where γ is the set of points at which $g'_k(x)$ violates the object-domain constraints and where g_k, G'_k and ϕ_k are estimates of f, F and the phase ψ of F, respectively. The algorithm is typically started by using an array of random numbers for $g_0(x)$ or for

$\phi_0(u)$. Figure 1 shows a block diagram of the iterative transform algorithm.

For the astronomy problem, the object-domain constraints are the object's nonnegativity and a (usually loose) support constraint. The diameter of the object can be computed since it is just half the diameter of the autocorrelation; however, the exact support of the object in general cannot be determined uniquely from the support of the autocorrelation [20], and so the support constraint cannot be applied tightly. For other problems, one may not have a nonnegativity constraint, but have a priori knowledge of a tighter support constraint [21].

For the problem of phase retrieval from two intensity measurements, $g'_k(x) = |g'_k(x)| \exp [i\theta'_k(x)]$ is complex-valued, and Step 4 becomes

$$g_{k+1}(x) = |f(x)| \exp [i\theta_{k+1}(x)] = |f(x)| \exp [i\theta'_k(x)], \qquad (8)$$

where $|f(x)|$ is the known modulus of the complex-valued object and θ_k is an estimate of the phase of the object. With this modulus constraint in the object domain, the error-reduction algorithm is precisely the Gerchberg-Saxton algorithm. In this paper we consider only the problem of phase retrieval from a single intensity measurement.

A measure of the convergence of the algorithm to a solution (a Fourier transform pair satisfying all the constraints in both domains) is the squared-error metric in the Fourier domain,

$$E_F^2 = N^{-2} \sum_u [\, |G(u)| - |F(u)| \,]^2, \qquad (9)$$

or in the object domain,

$$E_0^2 = \sum_{x \in \gamma} |g'_k(x)|^2 \qquad (10)$$

where γ is defined as in Eq. (7). Values of the error metrics mentioned later are the square roots of these expressions divided by $\sum_x |g'_k(x)|^2$, i.e., normalized root-mean-squared (NRMS) errors. It can be shown that the error-reduction algorithm converges in the sense that the squared error cannot increase with an increasing number of iterations [3].

Although it works well for the problem of phase retrieval from two intensity measurements, the error-reduction algorithm usually converges very slowly for the problem of phase retrieval from a single intensity measurement being considered here [3]. Several modifications of the iterative transform algorithm were made and tested, and most of them converged faster than the error-reduction algorithm [3]. To date, the most successful version is the hybrid input-output algorithm, which replaces Step 4 of the algorithm by [1, 3]

$$g_{k+1}(x) = \begin{cases} g'_k(x) & x \notin \gamma \\ g_k(x) - \beta g'_k(x), & x \in \gamma, \end{cases} \tag{11}$$

where β is a constant feedback parameter. Values of β between 0.5 and 1.0 work well. When using the hybrid input-output algorithm, $g_k(x)$ is no longer an estimate of $f(x)$; it is instead the input function used to drive the output $g'_k(x)$ [which is an estimate of $f(x)$] to satisfy the constraints. Hence only the object-domain error E_o is meaningful [3]. When using the hybrid input-output algorithm, even E_o does not always correlate with image quality as well as one would like. For this reason one may prefer to perform a number of cycles of iterations, where one cycle consists of, say, 20 to 50 iterations of the hybrid input-output algorithm followed by 5 to 10 iterations of the error-reduction algorithm, and note E_o only at the end of a cycle [3].

For a more complete description of the iterative algorithm, see Reference 3. Additional details concerning the implementation of the algorithm are given in Section 6. A description of the algorithm as it applies to a number of different problems is given in Reference 18.

3. METHOD FOR OVERCOMING SIMULTANEOUS TWIN IMAGES

Figure 2(a) shows a real, nonnegative object, $f(x)$, which has centro-symmetric (square) support, and Figure 2(b) shows the conjugate or twin image, $f^*(-x)$, which has the same Fourier modulus, $|F(u)|$. Since the object is real-valued, $f^*(-x) = f(-x)$. Figure 2(c) shows the output image of the iterative transform algorithm after a few hundred iterations using both a nonnegativity constraint and a support constraint consisting of the actual (assumed known a priori) square support of the object. Upon close inspection of Figure 2(c), it is seen that features of both $f(x)$ and $f^*(-x)$ are present (to see it, it helps to turn the page upside down). Often a few additional cycles of iterations are all that is needed to converge to one or the other of the twin images. An example of this is the sequence of output images shown in Figure 6(c), 6(d), 6(e) and 6(f) of Reference 2. However in the case of Figure 2(c) (although with a very large number of further iterations it may be possible to move away from this output having both twin images) this output image represents a fairly stable condition of stagnation. Like the fabled donkey standing between two bales of hay who starved to death because he was unable to decide which to eat, the algorithm is not readily able to move farther from the features of either of the twin images and so it is also prevented from moving closer to one over the other.

We have devised a method for getting beyond this condition: the reduced area support constraint method which consists of the following steps.

(1) Replace the current, correct mask defining the support constraint with a temporary one that (a) covers only a subset of the correct support including at least one of its edges and (b) has no 180° rotational symmetry (is not centro-symmetric).

(2) Perform a few iterations with the temporary mask.

(3) Replace the temporary mask with the correct one and continue with the iterations.

The reduced-area support constraint method is illustrated by the example shown in Figure 3. Figure 3(a) shows the stagnated output image of Figure 2(c) used as the input image to the algorithm. It had an error metric $E_0 = 0.027$. The correct support is a square. Figure 3(b) shows the reduced-area temporary support mask used for ten error-reduction iterations. Figure 3(c) shows the output image after the ten iterations. The correct square support constraint was then reinstated. Figure 3(d) shows the output image after 10 more iterations of the error-reduction algorithm ($E_0 = 0.060$); Figure 3(e) shows the output after an additional 60 iterations of the hybrid input-output algorithm plus 5 iterations of error-reduction ($E_0 = 0.027$); and Figure 3(f) shows the output after an additional three cycles of 40 hybrid input-output plus 5 error-reduction iterations each ($E_0 = 0.018$).

The reduced-area support constraint tends to favor some of the features of either $f(x)$ or $f^*(-x)$ over the other. Since it employs an incorrect support constraint, it cannot converge to a solution. However, when the correct support constraint is reinstated, one of the two twin images may have a sufficiently large advantage over the other that the algorithm can then converge toward that image.

In the small number of trials in which it was tested, the reduced-area support constraint method worked in the majority of the cases tried, but it is by no means guaranteed to work. If an

application of the method does not relieve the problem of stagnation with both twin images present, then one might try another application of the method using a different reduced-area temporary support constraint mask. The method as it stands has not yet been optimized as to the form of the temporary mask or the number or type of iterations that should be performed with the temporary mask. Nevertheless, the method has been shown to be very promising as a solution to the problem of stagnation with features of both twin images present.

For the example shown, knowledge that the simultaneous twin image mode of stagnation was present was obtained by visual inspection of the output image. The decision could also be automated by measuring the degree of symmetry of the image. An example of such a measure would be the ratio of the peak of the cross-correlation of $g'(x)$ with $g'^*(-x)$ to the peak of the autocorrelation of $g'(x)$.

4. METHODS OF OVERCOMING STRIPES

Several methods of overcoming the problem of stagnation associated with stripes across the image were attempted before successful methods were developed. Before describing the successful ones, two of the unsuccessful methods are mentioned here because they illustrate features of the problem.

4.1 SOME FEATURES OF STRIPES

The error metric, E_0 or E_F, can be considered as a function of an N^2-dimensional parameter space spanned by the values of $g(x)$ or of $\phi(u)$. Stagnation with stripes can be thought of as being stuck at a local minimum of the error metric. This local minimum is not very far from the global [$E_0 = E_F = 0$ for $g(x) = f(x)$] minimum, since the output image usually closely resembles the original object except for the presence of the stripes. It was thought that if the input image $g(x)$ were sufficiently perturbed then the estimate $g'(x)$

would be moved out of that local minimum and perhaps fall into the sought-after global minimum. Experimental tests were made in which increasing amounts of random noise were added to $g(x)$ and from these starting points more iterations were performed. It was found that even with very large amounts of added noise, in very few iterations the output image reverted back to the same point of stagnation having the same stripes as before. These experiments are an indication that the local minima characterized by stripes are very strong local minima.

A second unsuccessful method utilized the fact that since $g(x)$ is real-valued, and so $\phi(-u) = -\phi(u)$, the sinusoidal stripes must come from a conjugate pair of localized areas in the Fourier domain. In addition, since the iterative algorithm forces the output image to have the correct Fourier modulus at the sampled points, the error must also be a pure-phase error at the sampled points (see Section 4.4 for a discussion of the values between the samples). The spatial frequency of the stripes was measured to determine what area of the Fourier domain was in error. Constant phase terms were added to the Fourier transforms of the striped images in these areas in a conjugate-symmetric way, but the stripes remained in the image despite the use of a variety of constant phases and a variety of sizes of such areas. This was true despite the fact that, when similar constant phase errors were added to the Fourier transform of the object itself, the corresponding image had stripes that looked very much like the type of stripes that were produced by the mode of stagnation of the iterative algorithm; but these stripes went away after only one or two iterations, whereas the stripes produced by the stagnating iterative algorithm were very stable. This experience pointed out the spatial complexity of the phase error that caused the stripes.

4.2 VOTING METHOD

The key to solving the problem with the stripes is the fact that if the iterative algorithm is applied multiple times, each time with a different random starting guess, then the stripes of the various reconstructions will usually have different orientations and frequencies. This behavior was noted in Figure 9 of Reference 17. This implies that the errors will occur in different areas of the Fourier domain. Since the sinusoidal patterns usually become well defined, the areas of the phase errors are localized reasonably well. These features of the phase errors suggested a voting method. The idea is that if two of three Fourier phases are similar but a third is different, then the dissimilar phase is usually incorrect and should be discarded.

The voting method consists of the following steps:

(1) Generate three output images with different stripes by running the iterative transform algorithm three times, each with different random numbers for the initial input (if one of the output images is without stripes, then the problem is, of course, solved).

(2) Crosscorrelate the second and third images and their twins (the images rotated 180°) with the first image to determine their relative translations (to within a small fraction of a pixel, which can be accomplished by oversampling the crosscorrelation peak) and orientations.

(3) Fourier transform all three images.

(4) Subtract appropriate linear phase terms from the phases of the Fourier transforms of the second and third images, and conjugate if the orientation is opposite, to give the translations and orientations in the image domain that would cause the three images to be in registration. This

removes unwanted relative linear phase terms in the Fourier transforms (which could have been accomplished by translating the second and third images prior to Fourier transformation, but fractional-pixel translation is more easily accomplished in the Fourier domain).

(5) At each point u in the Fourier domain, compute the modulus of the difference between each pair of complex transforms. Average the two complex numbers that are the closest (discarding the third) and replace the complex value at that point with this average. Optionally: replace the modulus of the average with the measured modulus. Alternatively, one can take the phase midway between the two closest phases of the three, if proper attention is paid to modulo-2π questions.

(6) Inverse Fourier transform to yield the corresponding image.

The output of the voting method is used as the input to further iterations of the iterative transform algorithm.

If the method fails because two of the Fourier transforms have errors in the same location, then it should be repeated using different random numbers for the initial inputs to the iterative transform algorithm.

An example of the use of the voting method is shown in Figure 4. Figure 4(a) shows a diffraction-limited image of a satellite which is our object. It was formed from a digitized picture of a satellite within a 64 x 64 array imbedded in a 128 x 128 array. The digitized picture was low-pass filtered using the incoherent transfer function of a circular aperture of diameter 62 pixels (the Fourier transform of the object was multiplied by the autocorrelation of the circular aperture) to produce the object (a diffraction-limited image) shown in Figure 4(a). The sidelobes of the impulse response due to the circular aperture cause the diffraction-limited image to have a

small amount of energy well outside the support of the object, making the error metric $E_0 = 0.0026$ for this object. Because of this inherent slight inconsistency between the Fourier modulus data (which corresponds to the diffraction-limited image) and the support constraint, E_0 can never be driven to zero. Figures 4(b), 4(c) and 4(d) show three output images from the iterative transform algorithm, each generated using different random numbers for the starting input. The support constraint used was the 64 x 64 square support. Stripes of different spatial frequencies are clearly seen in each of the images. E_0 for the three output images is 0.0155, 0.0316 and 0.0038, respectively. For the output image shown in Figure 4(d), detection of the existence of the stripes is more difficult because of the low value of E_0, but inspection of an overexposed version of it clearly reveals them in the area outside the support of the object. Figure 4(e) shows the output of the voting method, for which $E_0 = 0.0680$, and Figure 4(f) shows the result of further iterations of the iterative transform algorithm, for which $E_0 = 0.0035$ and the stripes are successfully removed.

An advantage of the voting method is that one need not understand the nature of the phase error except that it is localized in different areas of the Fourier domain for different output images. The voting method may therefore be useful for other types of phase errors in addition to those characterized by stripes in the image.

4.3 PATCHING METHOD

The patching method, like the voting method, utilizes the fact that output images coming from different starting inputs usually have phase errors localized in different areas of the Fourier domain. The patching method uses an additional piece of information: since the stripes extend beyond the known support of the object (although they are dimmer there) they can be isolated and analyzed to determine approximately what area in the Fourier domain contains the localized

phase errors. With this information one can patch together a Fourier transform having less errors from two Fourier transforms which have these localized phase errors.

The patching method consists of the following steps:

(1-4) Perform steps (1) to (4) of the voting algorithm but using only two output images rather than three (if one of the output images is without stripes, then the problem is, of course, solved).

(5) For each of the two images, zero out the image in its support region. This isolates the stripes. Use a smooth apodization to avoid sidelobe problems in the Fourier domain.

(6) Fourier transform the isolated stripes from each.

(7) Smooth and threshold the Fourier modulus (after zeroing out a region about the origin to eliminate an undesired DC component) to generate a Fourier mask for each of the two images. These masks define the areas in the Fourier domain that have the phase errors.

(8) If the two masks overlap, repeat Step (7) using a larger threshold value or a smaller smoothing kernel or redo Steps (1) to (7) using another random input to start the iterative transform algorithm.

(9) Form a new Fourier transform having the phase of the Fourier transform of the first image except within its Fourier mask, where the phase of the Fourier transform of the second image is substituted.

(10) Inverse Fourier transform to yield the corresponding image.

The output of the patching method is used as the input to further iterations of the iterative transform algorithm.

An example of the use of the patching method is shown in Figures 5-7. Figure 5(a) shows the object, a diffraction-limited image of a satellite (E_0 = 0.0024), and Figures 5(b) (E_0 = 0.0268) and 5(c) (E_0 = 0.0503) show two output images from the iterative transform algorithm, each generated using different random numbers for the starting input. Upon close inspection, stripes of different spatial frequencies can be seen in each of the output images. Figure 6 shows the same thing as Figure 5, only heavily overexposed in order that the stripes over the object and beyond the support of the object can be more readily seen. Figure 7(a) shows the apodized mask used in the image domain which, when multiplied with the striped image, isolates the stripes. The resulting isolated stripes are shown in Figure 7(b). (A bias was added in the display of this result making the most negative value black, zero value grey and the largest value white.) Figure 7(c) shows the modulus of the Fourier transform of the isolated stripes, and Figure 7(d) shows the Fourier mask obtained by thresholding that Fourier modulus at a value 0.9 times the peak and smoothing with a 16 by 16 kernel. Figures 7(e) to 7(g) show the same things as Figures 7(b) to 7(d) but for the second striped image. The output of the patching method is shown in Figures 5(d) and 6(d)--the stripes in the two images were eliminated. Its error metric is E_0 = 0.00576 which is much lower than that for the striped images.

The voting and patching methods are both completely automated once it is decided that the iterative transform algorithm is stagnating on an image that has stripes. For the examples shown, knowledge that the striped-image mode of stagnation is present was obtained by visual inspection of the output image, from which it is quite obvious. This decision could also be automated, for example, by performing, for a given single output image, Steps (5) and (6) of the patching method and detecting the presence of especially bright areas in the Fourier domain.

4.4 ZERO REVERSAL OF THE FOURIER TRANSFORM

Comparison of the Fourier phase of the striped image with that of the original object yielded interesting insights into the properties of Fourier transforms.

Figure 8(a) shows the phase of the Fourier transform of the original object [shown in Figure 4(a)] and Figure 8(b) shows the upsampled phase of the area in Figure 8(a) outlined by the white square. Figure 8(c) and 8(d) show the same thing for the striped image [shown in Figure 4(b)]. To reduce the linear phase component, the centroid of the object was translated to the origin prior to Fourier transformation, and the striped image was translated to be in register with the object. The large circular pattern in Figure 8(a) is due to the simulation of the effects of diffraction by the circular aperture mentioned earlier. Outside the circle the Fourier transform has small nonzero values due to round-off error in the computer. Note that to upsample the phase, one must compute the phase of an upsampled complex Fourier transform which can in turn be computed by Fourier transformation of the object (or image) imbedded in a larger array padded with zeros.

Of particular interest is the phase within the four small squares drawn on Figures 8(b) and 8(d). In Figure 8(b), the phase within the upper right square "wraps around" one point, u_0, in the Fourier domain. That is, if one starts at a point u near u_0, as one progresses full circle around u_0 the phase slips by 2π radians. It is easily shown that this branch cut in the phase indicates that the Fourier modulus goes through a zero at u_0 [22]. A second-order zero [where $F(u_0) = 0$ has zero first partial derivatives as well] might not exhibit phase wrap-around, but they are rare compared to first-order zeros. The existence of zeros in F(u) implies an inherent $2\pi n$ (where n is an integer) ambiguity in the phase. Self-consistent phase unwrapping cannot logically be performed in such

cases. This is the usual case for Fourier transforms of images. From the presence or absence of phase wrap-around, it is evident that the Fourier transform goes through first-order zeros both within the upper right and lower left squares, but not within the two squares in the middle of Figure 8(b). On the other hand, for the case of the striped image, the presence or absence of first-order zeros is just the opposite, as can be seen in Figure 8(d). Where there are two first-order zeros for the original object there are not any for the striped image and vice versa. That is, the zeros are "reversed". (This zero reversal should not be confused with the flipping of complex zeros that is encountered in the analysis of uniqueness.) Also, by inspecting upsampled versions of the Fourier transforms, we found that the first-order zeros did not become higher-order zeros (which could cause the disappearance of the phase wrap-around), but they truly became nonzero. The difference between having and not having first-order zeros is extremely important: around a first-order zero the phase wraps around and varies very rapidly whereas otherwise the phase is relatively smooth. Note that the transitions from white (π phase) to dark ($-\pi$ phase) in Figure 8 are not jumps in phase per se; they are just an artifact of our ability to compute and display phase only modulo 2π.

If one draws a quadrilateral having vertices at the four points at which the zeros are reversed, one finds that the Fourier phase of the striped image differs from that of the object only (approximately) within the quadrilateral. Note that the phases outside the quadrilateral are practically the same in Figure 8(d) as in Figure 8(b). That is, the Fourier phase error for the striped image is localized in the area between the reversed zeros.

It is not accidental that the reversed zeros come in pairs. In order for the phase to be consistent in the surrounding area, a continuous path around the entire area of the localized phase error for the striped image must contain the same number of first-order zeros as for the Fourier transform of the object.

Initially it seems contradictory that the zeros of the Fourier transform of the striped output image could be different from those of object's Fourier transform, since the striped image and the object have exactly the same Fourier modulus at the sampled points. However, this possibility arises since we are dealing with sampled data in the computer. The zeros only rarely fall on the sampling lattice: they usually fall some distance between the samples. In the presence of even the slightest error, including round-off error due to the finite word length used by the computer, it becomes difficult to see, even from a heavily oversampled Fourier modulus, whether it goes through zero or merely comes very close to it. More important, though, is that since the striped image has energy throughout image space, rather than being confined to the support of the object, its Fourier modulus is aliased and differs from the Fourier modulus of the object for points off the sampling lattice. Hence its Fourier transform can truly have zeros where the object's Fourier transform does not, and vice versa, despite their having the same Fourier modulus at the sampled points.

4.5 LINES OF REAL AND IMAGINARY ZEROS

Figure 9 shows the lines where the real and imaginary parts of the Fourier transform of the object (in this case translated to be in one quadrant of the array) are zero, for the same area of the Fourier domain shown in Figure 8(b). We will refer to these lines as the lines of real zeros and lines of imaginary zeros. (Note we are referring here to the zeros in the two-dimensional real plane, not to the zeros in the complex plane which are frequently discussed in regard to the uniqueness of phase retrieval.) The lines of real zeros were computed by scanning across each line and each column of an oversampled version of the real part of the Fourier transform and noting where it changed sign. The lines of imaginary zeros were found in a similar manner. In Figure 9 the real zeros are denoted

by dark lines and the imaginary zeros by light lines on a grey background. The complex Fourier transform goes through a zero wherever both the real part and the imaginary part are zero, that is, where the dark lines and light lines intersect. The entire discussion regarding the zeros of the Fourier transform of the striped image versus those of the object can be explained in terms of Figure 9 and a similar picture for the striped image case, as well as by using Figure 8.

Except for special cases, in Figure 9 the lines of real zeros and imaginary zeros cross at single points rather than being tangent to one another over extended intervals; hence the zeros tend to occur at discrete points.

In addition, notice that the lines of imaginary zeros have a strong tendency to be half-way between two lines of real zeros, and vice versa. This can be understood as follows. Half-way between two neighboring lines of real zeros (think of them as a single-level topographic map of the real part) one would expect to find a line of local maxima or minima, like ridges or gullies, respectively. These ridges (or gullies) have the property that they are local maxima (minima) along all directions except along the length of the ridge (gully). Since the object $f(x, y)$ was translated to the upper left quadrant of the plane [i.e., $f(x, y) = 0$ for $x < 0$ and for $y < 0$ making it "causal"] then $F(u, v)$ satisfies the Hilbert transform relationships (see Appendix A):

$$F_I(u, v) = \frac{1}{\pi} P \int \frac{F_R(u, v')}{v' - v} dv' \qquad (12)$$

$$= \frac{1}{\pi} P \int \frac{F_R(u', v)}{u' - u} du' \qquad (13)$$

where $F = F_R + iF_I$, F_R and F_I being real-valued and P denotes the Cauchy principal value. For clarity in this discussion and in

Appendix A, we use (x, y) as the two-dimensional coordinates in object space and (u, v) in Fourier space rather than the vector notation used elsewhere in this paper. Since $1/(v' - v)$ is much larger near $v' = v$ than elsewhere, one would expect the integrand near $v' = v$ to dominate the integral of Eq. (12). If the point (u_o, v_o) is at a ridge (or a gully) of $F_R(u, v)$, then one would expect $F_R(u_o, v')$ to have a local maximum (minimum) at $v' = v_o$ and be closely approximated by a quadratic in a small region of v' centered about v_o, since $F_R(u_o, v')$ is a bandlimited function of v' (it is the Fourier transform of a function of finite extent). Therefore, since the numerator $F_R(u_o\ v')$ is even about $v' = v_o$ and the denominator $(v' - v_o)$ is odd about $v' = v_o$, the integrand is odd about $v' = v_o$ and the contribution to the integral from the neighborhood about v_o is near zero. Since that neighborhood is the part of the integral that usually dominates, it is easily seen why $F_I(u, v)$ tends to be zero near the ridges and gullies of $F_R(u, v)$. The same can be shown from Eq. (13). The same argument can be used to show why $F_R(u, v)$ tends to be zero near the ridges and gullies of $F_I(u, v)$.

5. METHOD FOR OVERCOMING TRANSLATED SUPPORT

Since $f(x - x_o)$ has the same Fourier modulus as $f(x)$, the location of the object's support is arbitrary. Frequently the image partially reconstructed by the algorithm will not be in perfect registration with the support constraint. Then enforcing the support constraint causes an inadvertent truncation of part of the desired image, causing the algorithm to stagnate. In addition to the enlarging support method described in Section 6, a method of combatting this stagnation problem is to dynamically translate either the support constraint or the image.

The amount of translation to be used can be determined as follows. Compute the total energy of the output image, $g'_k(x)$,

(i.e., square and sum) over the area of the support constraint for the current position of the support constraint and for the support constraint translated by one or two pixels in every direction. The support constraint should be translated to the position for which the energy is maximized. This can be done occassionally or at every iteration. Alternatively, compute the cross-correlation of the support mask with $g'_k(x)$ or with $|g'_k(x)|^2$ and translate according to the peak of the cross-correlation. This method would be particularly effective if, just prior to performing it, a support constraint larger than the usual support were used for a few iterations; that would give the truncated part of the image a chance to establish itself.

6. ITERATIVE TRANSFORM ALGORITHM DETAILS

In the past, some researchers have had varying success in applying the iterative transform algorithm to phase retrieval from a single intensity measurement. In this section, a number of additional aspects of making the iterative algorithm work are given as an aid to the practical implementation of the algorithm.

Recall from Section 2 that the heart of the algorithm consists of several cycles of iterations, where one cycle consists of K_1 iterations of the hybrid input-output algorithm [Eqs. (4), (5), (6) and (11)] followed by K_2 iterations of the error-reduction algorithm [Eqs. (4)-(7)]. Our experience has shown that values of K_1 from 20 to 100, of K_2 from 5 to 10, and of the feedback parameter β from 0.5 to 1.0 (use, say, 0.7) work well.

The discrete Fourier transforms are computed using the Fast Fourier Transform (FFT) algorithm. The sampling in the Fourier domain should be fine enough to ensure that the object domain array size is at least twice the width and height of the object itself, which is equivalent to achieving the Nyquist sampling rate for $|F(u)|^2$.

A straightforward method to evaluate Eq. (5) is to compute the phase from the real and imaginary parts of $G_k(u)$, then combine it with $F(u)$ to form $G'_k(u)$, and finally compute the real and imaginary parts of $G'_k(u)$ (which are required by the FFT) from its modulus and phase. Alternatively, one can employ

$$G'_k(u) = G_k(u)\ |F(u)| \ /\ [\,|G_k(u)| + \delta\,] \tag{14}$$

where δ is a very small number used to prevent overflow problems in the rare event that $G_k(u) = 0$. (For some computers one can use $\delta = 0$ with no ill effects.)

The data one must have available is an estimate, $|\hat{F}(u)|$, of the modulus, $|F(u)|$, of the Fourier transform of the object. Although the iterative transform reconstruction algorithm is not hypersensitive to noise, care must be taken to obtain the best possible estimate of the Fourier modulus, which may involve considerable compensation of the raw data [23], depending on how it is collected. In many circumstances one can estimate the expected value of the normalized root-mean-squared (NRMS) error of the data:

$$E_{|\hat{F}|} = \left[\frac{\sum_u [\,|\hat{F}(u)| - |F(u)|\,]^2}{\sum_u |F(u)|^2} \right]^{1/2} . \tag{15}$$

As described later, this is useful for deciding when one is close enough to a solution.

For the astronomy problem one has a nonnegativity constraint in the object domain. Furthermore, one can compute upper bounds on the support of the object in any of several ways [20]. The simplest way is to use a rectangle that is half the size, in each dimension, of the smallest rectangle that encloses the autocorrelation, which is given by the inverse Fourier transform of $|F(u)|^2$. If the actual support of the object is known a priori, then that should of course

be used. Any other types of a priori information should be used during the iterations if available. The support constraint can in general be defined by a mask which is unity within the support and zero outside. If the support constraint and the set of points are defined as binary mask arrays; then the computations of Eqs. (7), (10) and (11) can be performed arithmetically without the use of logic, which is advantageous when using array processors.

There are many ways to pick an initial input to the algorithm. Although claims have been made that a certain crude estimation of the phase offers a superior starting point [24], others have found that random numbers do as well or better [25]. Having an initial input close to the true solution reduces the number of iterations required and might help to avoid some of the stagnation problems. If another reconstruction method (Knox-Thompson [26] for astronomical speckle interferometry for example) has yielded an image, then that image would be the appropriate starting input. One can either view the other reconstruction method as a means for supplying starting inputs for the iterative transform algorithm or view the iterative transform algorithm as a means for "cleaning up" images reconstructed by the other method. If no other initial estimate for the object is available, then one should use random numbers in the object domain or for the Fourier phase, giving an unbiased start to the algorithm. In the object domain, a convenient starting guess, $g_0(x)$, can be formed by filling the support mask with random numbers. Another method [3] is to threshold the autocorrelation (at say, 0.005 its maximum value), demagnify that by a factor of two in each dimension by discarding every other row and every other column, and finally fill the resulting shape with random numbers. (Note that this shape does not necessarily contain the support of the object [20].)

The algorithm can be made to converge faster and avoid a stagnation problem (see Section 5) if the support mask is chosen to be somewhat smaller than the correct support for the first cycle or two

of iterations. Since it is the incorrect support, the smaller support mask is inconsistent with Fourier modulus and stagnation will eventually occur when it is used. Nevertheless, the smaller support mask helps to force most of the energy of the output, g'(x), into a confined region in fewer iterations. After this has happened the support mask should be enlarged to the correct support constraint for the object. This enlargement of the mask could be done in more than one step if desired. When the algorithm has nearly finished reconstructing the object, it is often beneficial to make the support mask even larger than the correct support for the object. This helps to ensure that no parts of the object are being inadvertently truncated by the support constraint. The progression from a smaller support mask to a large one also helps to avoid having edges of the output image biased toward falling right at the edges of the support mask. Use of the method described in Section 5 does this as well and is recommended for use whenever truncation is suspected (or to be safe at the end of each cycle or even after every iteration).

As the iterations progress, the normalized RMS (NRMS) error in the object domain,

$$E_o = \left[\frac{\sum_{x \epsilon \gamma} |g'_k(x)|^2}{\sum_{x} |g'_k(x)|^2} \right]^{1/2} \qquad (16)$$

[compare with Eq. (10)] should be computed. The NRMS error is a measure of how close the current Fourier transform pair is to a solution. Note that the denominator of the above error metric is a constant that need be computed only once. Note also that at the end of a cycle $E_o \approx E_F$, the NRMS error in the Fourier domain [3]. When E_o goes significantly below $E_{|\hat{F}|}$ of Eq. (15), one has a solution consistent with the measured data and constraints to within the limits of the error in the given data. It is unlikely that E_o

will ever go to zero because noise in $|\hat{F}(u)|$ almost always results in a Fourier modulus that is inconsistent with either the nonnegativity constraint or any reasonable support constraint, or both. This can be seen from the fact that an autocorrelation computed from a noisy $|\hat{F}(u)|$ will ordinarily have negative values at some points (which could only arise from an object having negative values) and will ordinarily have (possibly small) nonzero values far beyond the extent of the true autocorrelation of the object. This problem can be alleviated by setting equal to zero the values of the autocorrelation that are negative or lie beyond some assumed autocorrelation support; but even then, noise remains and there will be no nonnegative image completely consistent with the Fourier modulus estimate. On the other hand, in the presence of noise there will ordinarily exist an output image $g'(x)$ that is in better agreement with the noisy data than the true image is. Consequently, for the case of noisy data, a "solution" is not found until E_0 is decreased to a level somewhat less than $E_{|F|}$. An exception to this is for the very low noise case in which the dominant error is sidelobe energy which spills outside the object's support due to diffraction effects, as for the examples shown in Section 4.

If all goes well, the iterative transform algorithm will converge to a solution after a small number of cycles of iterations. If there are multiple solutions, the iterative transform algorithm is capable of finding any one of them, depending on the starting input [27, 28]. Confidence that the solution is the one and only true solution can be increased by performing two or more trials of the algorithm, each time using different random numbers for the initial input.

In some cases the iterative transform algorithm will stagnate before reaching a solution. The algorithm can be considered to have stagnated if the error E_0 has failed to decrease after three additional cycles. While some objects can be reconstruced very easily, requiring only one or two cycles, other more difficult objects can

require many cycles comprising well over a thousand iterations. Consequently, one should not too readily jump to the conclusion that the algorithm has stagnated. It often occurs that very slow progress is made for many iterations, but then the algorithm suddenly finds its way and rapid progress is made in just a few iterations.

If the iterative transform algorithm does stagnate, then one can start over with a different set of random numbers for the initial input; alternatively, sever methods for getting out of the stagnated condition are possible. Sometimes changing the support constraint (enlarging it or translating it as described in Section 5) is what is required. The need for doing this can be established by following the steps suggested at the end of Section 5. Sometimes altering the feedback parameter, β, helps. Temporarily using a larger value for β, say 1.2, causes larger changes to be made and may move the output away from the condition of stagnation; but this should not be carried on for too many iterations since it causes the algorithm to become unstable. If the support mask is centrosymmetric or nearly so, then the simultaneous twin images can be present. This condition can be detected visually (comparing the output image with a second version of it rotated 180° helps) or by the method suggested at the end of Section 3. If this condition is suspected, then use the method of overcoming the problem of simultaneous twin images described in Section 3. Even if the twin-image problem is not present, the method might move the output image out of the condition of stagnation. The mode of stagnation characterized by stripes is easily detected by looking for stripes outside the support region in a picture of $g'_k(x)$ that is heavily overexposed; alternatively one could use the method described at the end of Section 4.3. If stripes are present, the iterations should be continued until stagnation occurs, since (1) the stripes may go away naturally and (2) further iterations cause the stripes to become more nearly sinusoidal which is equivalent to the phase errors being

confined to a smaller, more distinct area of the Fourier domain, which makes them easier to overcome by the methods described in Section 4.

Other tricks may be helpful or necessary for certain special cases. For example, if the object consists of some interesting details superimposed on a diffuse background, then the defogging method can make the reconstruction of the object easier [24].

7. SUMMARY AND CONCLUSIONS

In many cases of interest, the problem of phase retrival from a single intensity measurement can be solved by a straightforward application of a few cycles of the iterative transform algorithm. For some cases, the algorithm stagnates before reaching a solution consistent with the data and constraints. Three different modes or conditions of stagnation have been identified: simultaneous twin images, stripes superimposed on the image, and unintentional truncation by the support constraint. Methods for overcoming each of these modes of stagnation have been devised and have been demonstrated to be effective for particular examples. The use of these methods in conjunction with the iterative transform algorithm greatly enlarges the class of objects which can be reconstructed successfully. This has also helped to provide further empirical evidence of the uniqueness of the solution for two-dimensional objects. Some previous doubts of uniqueness, tied to an inability of the algorithm to converge in some instances, have been removed [28]. In particular, we have definitively shown that the striped images represent a local minimum rather than a true ambiguity.

In the course of investigating the stripes phenomenon, insight was gained into some of the properties of the Fourier transforms of images. The Fourier transform of a striped image has a phase that differs from that of the Fourier transform of the object in a fairly

well defined region that is determined by the locations of zeros of the Fourier transform which are reversed, i.e., where first-order zeros appear or disappear. First-order zeros are common in the Fourier transforms of images. Attempts at phase unwrapping, as required by the Knox-Thompson method, utilizing multiple paths of integration will fail unless proper attention is paid to the branch cuts associated with first-order zeros. If an image is causal, then the lines of real and imaginary zeros of its Fourier transform follow along the ridges and gullies of the imaginary part and the real part, respectively, which can be understood from the Hilbert transform relationship.

APPENDIX A
HILBERT TRANFORMS FOR TWO DIMENSIONS

Let $f(x, y)$ be zero for $x < 0$ and for $y < 0$ (Let all integrations be understood to be from $-\infty$ to $+\infty$).

$$F(u, v) = F_R(u, v) + iF_I(u, v) \tag{A1}$$

$$= \iint f(x, y) \exp [-i2\pi(ux + vy)] \, dxdy$$

$$= \int \left[\int f(x, y) \exp (-i2\pi ux) \, dx \right] \exp (-i2\pi vy) \, dy$$

$$\equiv \int \tilde{f}(u, y) \exp (-i2\pi vy) \, dy, \tag{A2}$$

where $\tilde{f}(u, y)$ is zero for $y < 0$. Fixing u to be a constant for the moment, $F(u, v)$ is the one-dimensional Fourier transform of $\tilde{f}(u, y)$, which is zero for $y < 0$ (i.e., it is causal in y), in which case we have the Hilbert transform relationships [29]

$$F_I(u, v) = \frac{1}{\pi} P \int \frac{F_R(u, v')}{v' - v} \, dv' \tag{A3}$$

and

$$F_R(u, v) = -\frac{1}{\pi} P \int \frac{F_I(u, v')}{v' - v} \, dv' \tag{A4}$$

where P denotes the Cauchy principal value. This is true for all values of (u, v). By a similar argument we have

$$F_I(u, v) = \frac{1}{\pi} P \int \frac{F_R(u', v)}{u' - u} \, du' \tag{A5}$$

and

$$F_R(u, v) = -\frac{1}{\pi} P \int \frac{F_I(u', v)}{u' - u} \, du'. \tag{A6}$$

ACKNOWLEDGEMENT

This research was supported by the U.S. Air Force Office of Scientific Research under Contract F49620-82-K-0018.

Portions of this paper were presented at the October 1984 Annual Meeting of the Optical Society of America in San Diego [30].

REFERENCES

1. J.R. Fienup, "Reconstruction of an Object from the Modulus of Its Fourier Transform," Opt. Lett. 3, 27-29 (1978).

2. J.R. Fienup, "Space Object Imaging Through the Turbulent Atmosphere," Opt. Eng. 18, 529-534 (1979).

3. J.R. Fienup, "Phase Retrieval Algorithms: A Comparison," Appl. Opt. 21, 2758-2769 (1982).

4. G.H. Stout and L.H. Jenson, X-Ray Structure Determination (Macmillan, London, 1968).

5. C.Y.C. Liu and A.W. Lohmann, "High Resolution Image Formation through the Turbulent Atmosphere," Opt. Commun. 8, 372-377 (1973).

6. P.J. Napier and R.H.T. Bates, "Inferring Phase Information from Modulus Information in Two-Dimensional Aperture Synthesis," Astron. Astrophys. Suppl. 15, 427-430 (1974).

7. B.R. Frieden and D.G. Currie, "On Unfolding the Autocorrelation Function," J. Opt. Soc. Am. 66, 1111A (1976).

8. J.E. Baldwin and P.J. Warner, "Phaseless Aperture Synthesis," Mon. Not. R. Astr. Soc. 182, 411-422 (1978).

9. R.H.T. Bates, et al., "Fourier Phase Problems Are Uniquely Solvable in More than One Dimension. I: Underlying Theory," Optik 61, 247-262 (1982); "...II: One-Dimensional Considerations," Optik 62, 131-142 (1982); "... III: Computational Examples for Two Dimension," Optik 62, 219-230 (1982).

10. H.H. Arsenault and K. Chalasinska-Macukow, "The Solution to the Phase Retrieval Problem Using the Sampling Theorem," Opt. Commun. 47, 380-386 (1983); K. Chalasinska-Macukow and H.H. Arsenault, "Fast Iterative Solution to Exact Equations for the Two-Dimensional Phase-Retrieval Problem," J. Opt. Soc. Am. A 2. 46-50 (1985).

11. A. Levi and H. Stark, "Image Restoration by the Method of Generalized Projections with Application to Restoration from Magnitude," J. Opt. Soc. Am. A 1, 932-943 (1984).

12. H.V. Deighton, M.S. Scivier and M.A. Fiddy, "Solution of the Two-Dimensional Phase Retrieval Problem," Opt. Lett. 10, 250-251 (1985).

13. R.W. Gerchberg and W.O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," Optik 35, 237-246 (1972).

14. R.W. Gerchberg, "Super-Resolution through Error Energy Reduction," Optica Acta 21, 709-720 (1974).

15. W.O. Saxton, Computer Techniques for Image Processing in Electron Microscopy (Academic Press, New York, 1978).

16. M.A. Fiddy, B.J. Brames, and J.C. Dainty, "Enforcing Irreducibility for Phase Retrieval in Two Dimensions," Opt. Lett. 8, 96-98 (1983).

17. G.B. Feldkamp and J.R. Fienup, "Noise Properties of Images Reconstructed from Fourier Modulus," in 1980 International Optical Computing Conference, W.T. Rhodes, ed., Proc. SPIE 231, 84-93 (1980).

18. J.R. Fienup, "Reconstruction and Synthesis Applications of an Iterative Algorithm," in Transformations in Optical Signal Processing, W.T. Rhodes, J.R. Fienup, and B.E.A. Saleh, eds., Proc. SPIE 373, 147-160 (1981).

19. D.C Youla, "Generalized Image Restoration by Method of Alternating Orthogonal Projections," IEEE Trans. Circuits and Systems CAS-25, 694-702 (1978).

20. J.R. Fienup, T.R. Crimmins, and W. Holsztynski, "Reconstruction of the Support of an Object from the Support of Its Autocorrelation," J. Opt. Soc. Am. 72, 610-624 (1982).

21. J.R. Fienup, "Phase Retrieval from a Single Intensity Distribution," in Optics in Modern Science and Technology, Conference Digest for ICO-13, August 20-24, 1984, Sapporo, Japan, pp. 606-609.

22. M.S. Scivier and M.A. Fiddy, "Phase Ambiguities and the Zeros of Multi-Dimensional Band-Limited Functions," J. Opt. Soc. Am. A 2, 693-697 (1985).

23. J.R. Fienup and G.B. Feldkamp, "Astronomical Imaging by Processing Stellar Speckle Interferometry Data," in Applications of Speckle Phenomena, W.H. Carter, ed., Proc. SPIE 243, 95-102 (1980).

24. R.H.T. Bates and W.R. Fright, "Composite Two-Dimensional Phase-Restoration Procedure," J. Opt. Soc. Am. 73, 358-365 (1983).

25. R.J. Sault, "The Fourier Phase Problem," FLEURS Internal Report No. 83/1, University of Sydney (1983).

26. K.T. Knox and B.J. Thompson, "Recovery of Images from Atmospherically Degraded Short-Exposure Photographs," Astrophys. J. Letters 193, L45-L48 (1974).

27. P. vanToorn, A.H. Greenaway and A.M.J. Huiser, "Phaseless Object Reconstruction," Optica Acta 7, 767-774 (1984).

28. J.R. Fienup, "Experimental Evidence of the Uniqueness of Phase Retrieval from Intensity Data," in Indirect Imaging, Proceedings of URSI/IAU Symposium, 30 Aug. to 2 Sept. 1983, Sydney, Australia, ed. J.A. Roberts (Cambridge University Press, Cambridge, 1984), pp. 99-109.

29. R.N. Bracewell, The Fourier Transform and Its Applications, Second Edition (McGraw-Hill, New York, 1978), pp. 271-272.

30. J.R. Fienup and C.C. Wackerman, "Improved Phase Retrieval Algorithm," J. Opt. Soc. Am. A 1, 1320 A (1984).

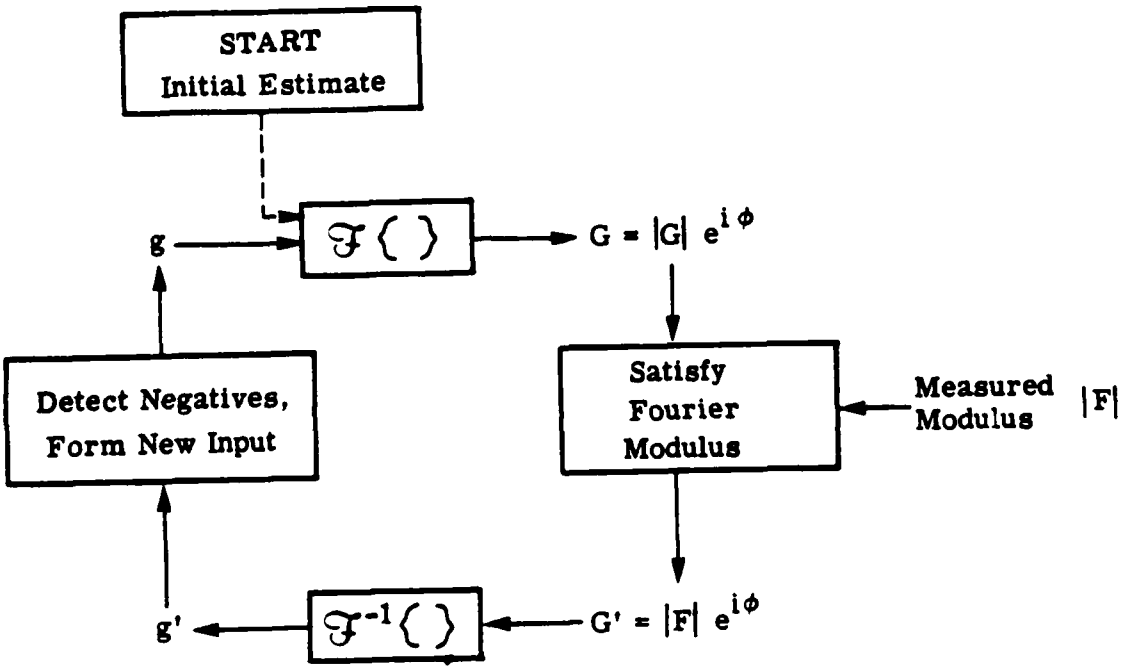


Figure 1. Block diagram of the iterative transform algorithm.

Figure 2. Simultaneous twin images problem. (a) Object f(x); (b) twin image f*(-x); (c) output image from the iterative transform algorithm which has stagnated with features of both.

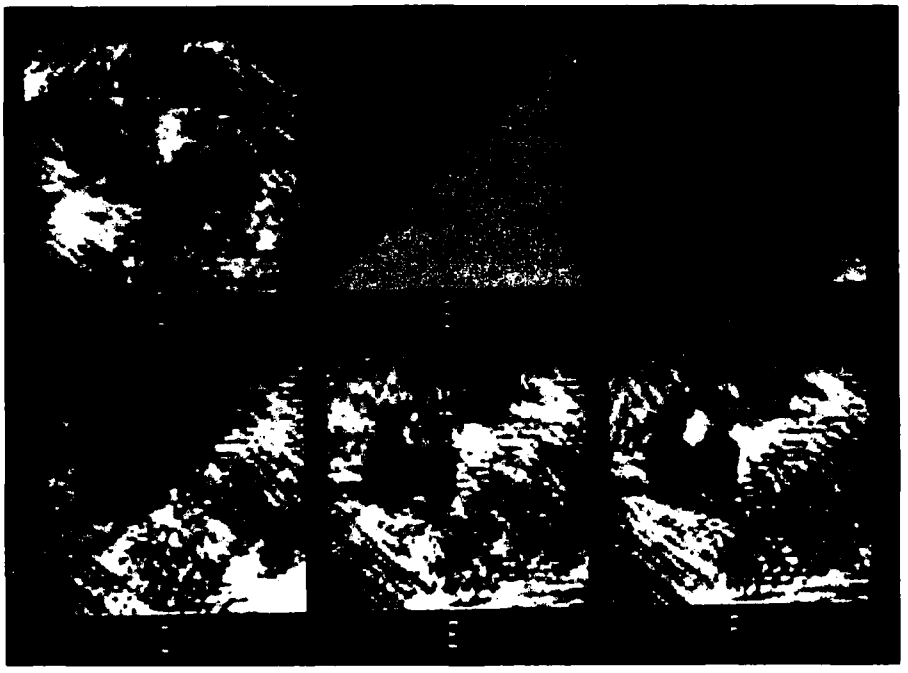


Figure 3. Reduced-area support constraint method for overcoming the problem of simultaneous twin images. (a) Stagnated output image; (b) mask defining the temporary reduced-area support constraint; (c) output image after ten iterations using temporary support; (d), (e) and (f) output images after further iterations using the correct support.

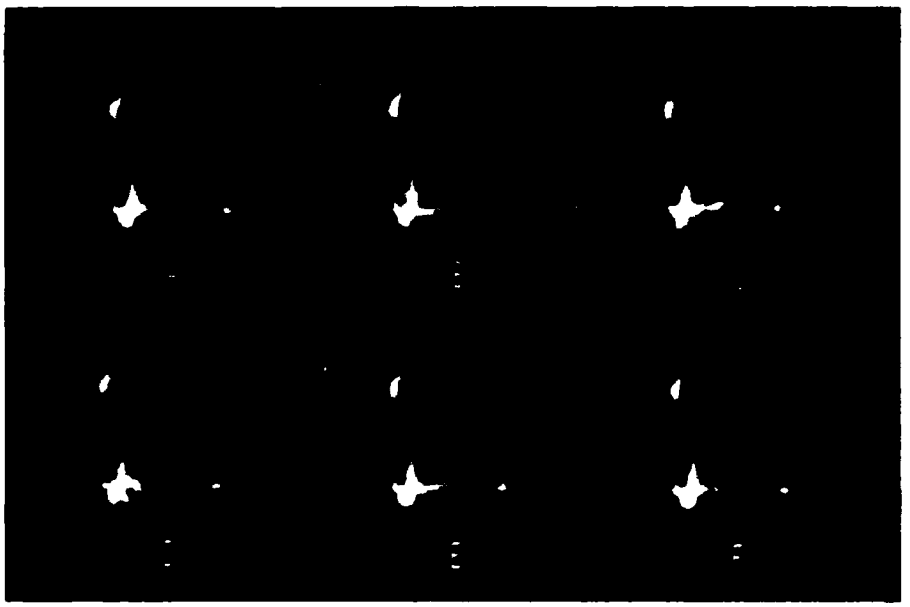


Figure 4. Voting method for eliminating stripes in the output image. (a) The object; (b), (c), and (d) output images from the iterative transform algorithm each with different stripes; (e) output of voting method; (f) output image after further iterations.

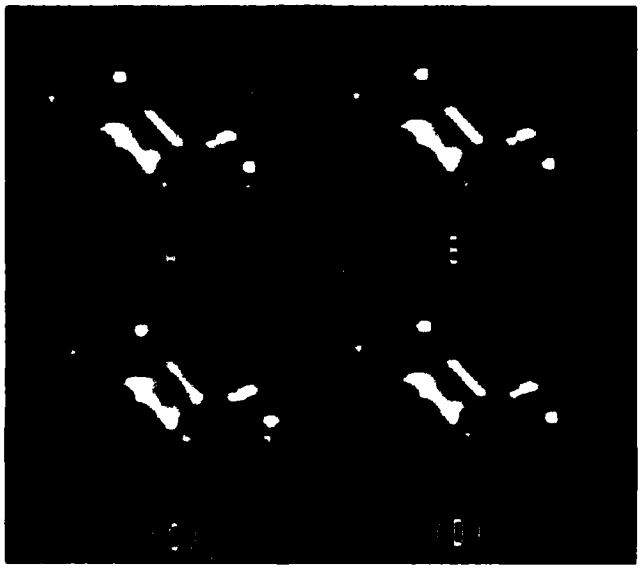


Figure 5. Patching method for eliminating stripes in the output image. (a) The object; (b) and (c) output images from the iterative transform algorithm, each with different stripes; (d) output of patching method.

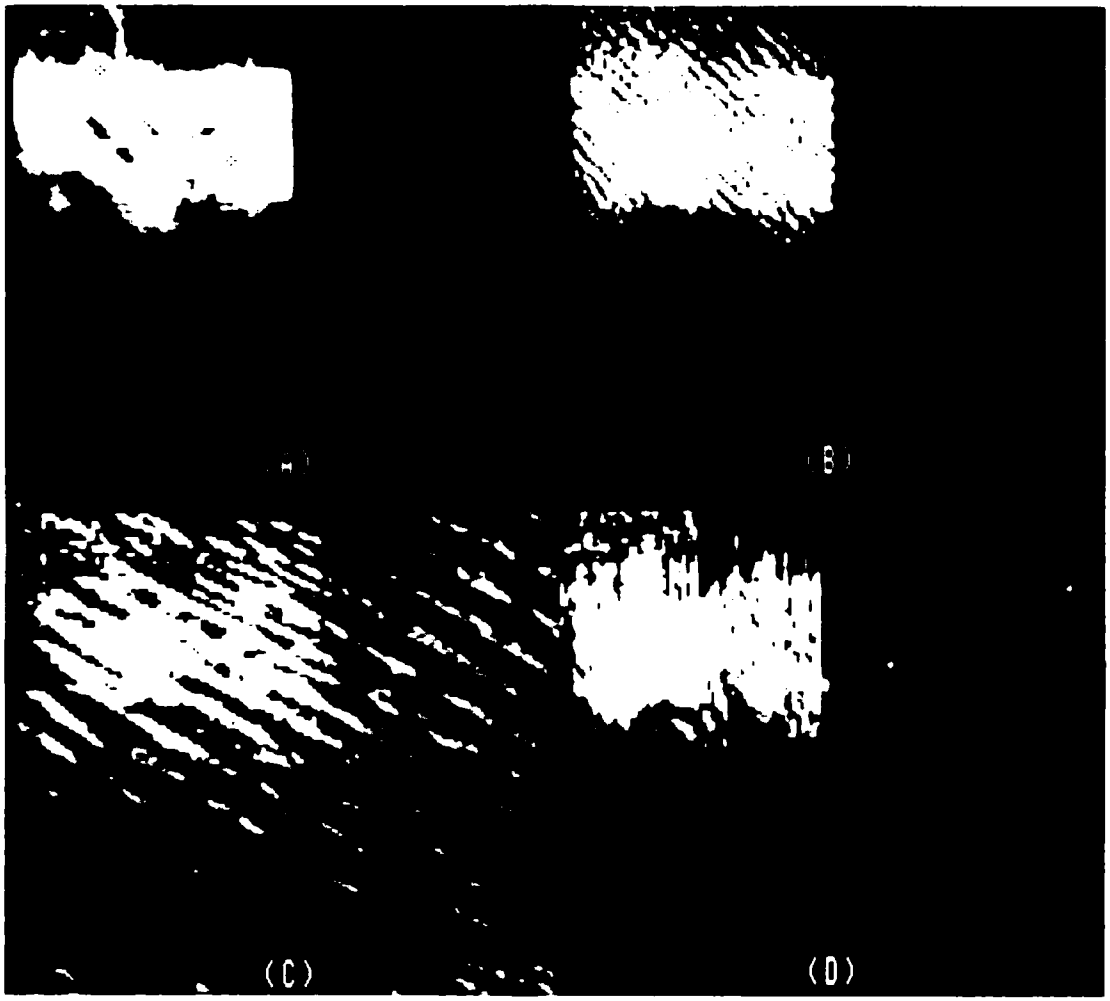


Figure 6. Same as Figure 5 but overexposed to emphasize the stripes.

Figure 7. Details of the patching method. (a) Mask used to isolate the stripes of the output images; (b) stripes isolated from the first image; (c) modulus of the Fourier transform of the isolated stripes from the first image; (d) Fourier mask obtained by thresholding and smoothing (c). (e) to (g) are the same as (b) to (d) but for the second image.

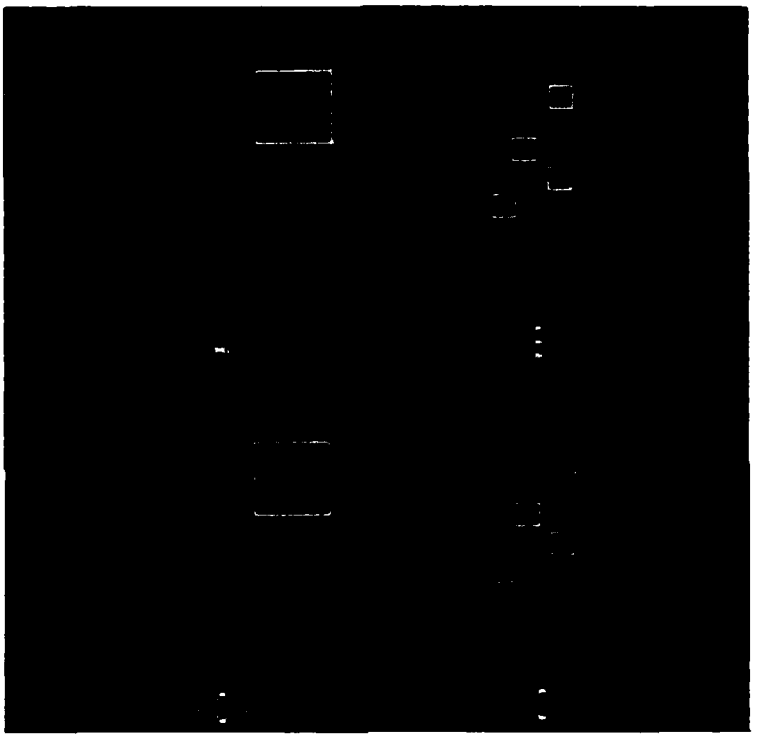


Figure 8. Fourier phases. (a) Fourier phase of the object; (b) upsamples phase from the area in (a) outlined by the square; (c) Fourier phase of the striped output image; (d) upsampled phase from the area in (c) outlined by the square. The (u, v) zeros of the complex Fourier transforms are reversed in the areas enclosed in squares in (b) and (d).

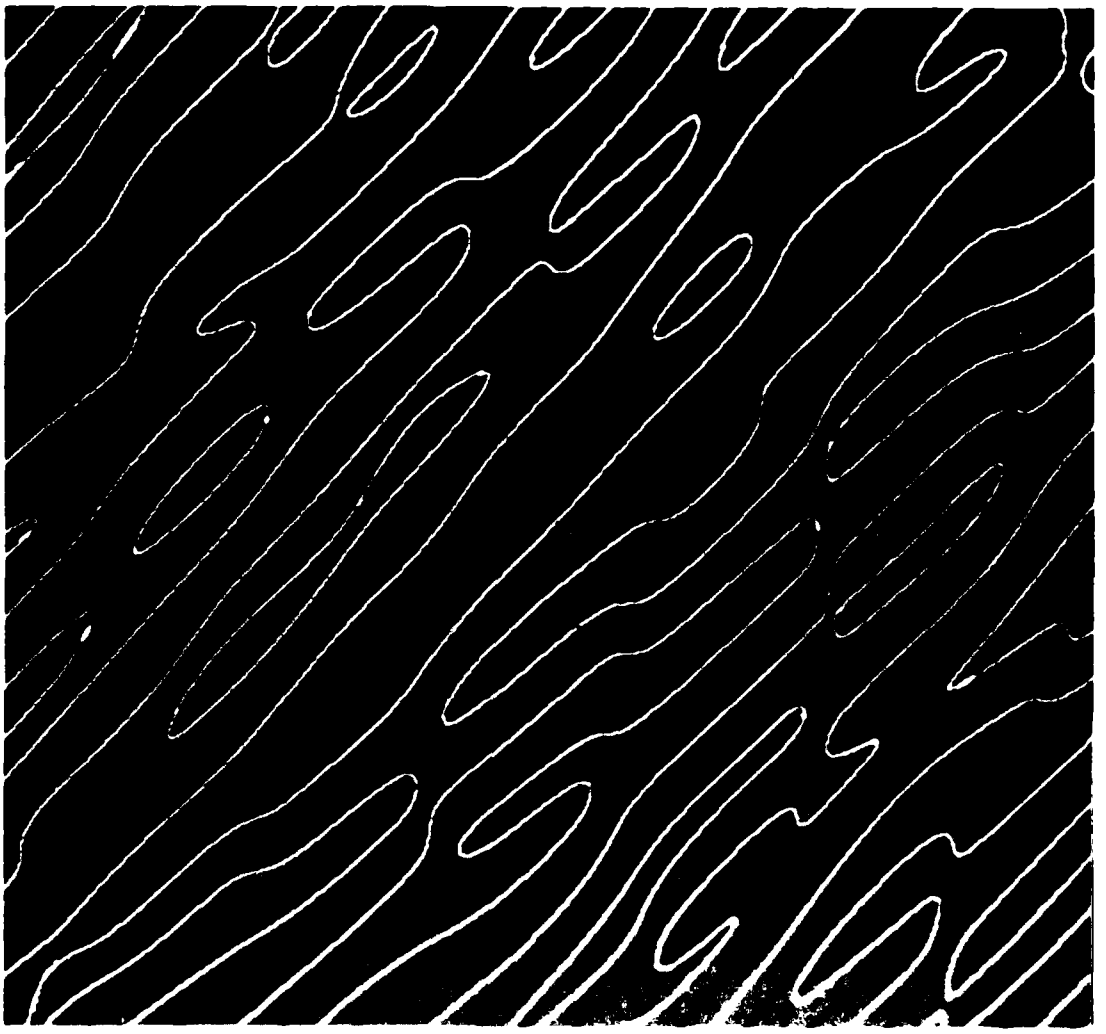


Figure 9. Locations of the zeros of the real part (dark lines) and the imaginary part (white lines) of the Fourier transform of the object. The object was translated to be causal and the area of its Fourier transform shown here is that shown in Figure 8(b).

Appendix D

UNIQUENESS OF 2-D PHASE RETRIEVAL SHOWN EMPIRICALLY

J.C. Fienup and C.C. Wackerman

Submitted to the Journal of the Optical Society A.

Appendix D

UNIQUENESS OF 2-D PHASE RETRIEVAL SHOWN EMPIRICALLY

by

J.R. Fienup and C.C. Wackerman

Environmental Research Institute of Michigan
P.O. Box 8618, Ann Arbor, MI 48107

Abstract

The uniqueness of the phase retrieval problem — the reconstruction of an object from the modulus of its Fourier transform — has been debated in recent years. Experimental reconstruction results are shown which support the theory that two-dimensional phase retrieval is usually unique. It is also shown that uniqueness in a practical sense is not destroyed by noise in the Fourier modulus data.

Submitted to the Journal of the Optical Society A.

1. INTRODUCTION

The phase retrieval problem considered in this paper is the reconstruction of an object function, $f(x, y)$, from the modulus, $|F(u, v)|$, of its Fourier transform

$$F(u, v) = |F(u, v)| \exp [i\Psi(u, v)] = \mathcal{F}[f(x, y)]$$

$$= \int\int_{-\infty}^{\infty} f(x, y) \exp [-i2\pi(ux + vy)] \, dx \, dy \, . \qquad (1)$$

It is equivalent to the reconstruction of the Fourier phase, $\Psi(u, v)$, from the Fourier modulus and to the reconstruction of $f(x, y)$ or $\Psi(u, v)$ from the autocorrelation function

$$r(x, y) = \mathcal{F}^{-1}[\, |F(u, v)|^2] \, . \qquad (2)$$

This problem arises in several disciplines including optical and radio astronomy, wavefront sensing, holography, and remote sensing.

There are the omnipresent ambiguity that the object, $f(x, y)$, any translation of the object, $f(x - x_0, y - y_0)$, the twin image, $f^*(-x - x_0, -y - y_0)$, and any of these multiplied by a constant of unity magnitude, $\exp [i\phi_c]$, all have exactly the same Fourier modulus. These ambiguities change only the object's position or orientation, not its appearance. If they are the only ambiguities, then we refer to the object as being unique. A solution is considered to be ambiguous only if it differs from the object in ways other than these omnipresent ambiguities.

If nothing is known about the object, then reconstruction from its Fourier modulus is generally ambiguous except for very special cases. Fortunately, for many applications one has additional a priori knowledge about or constraints on the object. In the astronomy application, for example, the object's spatial brightness distribution $f(x, y)$ is a real, nonnegative function. For several applications, one has a support constraint, i.e., the object is

known to be zero outside some finite area. Even if the support constraint is not known a priori, upper bounds can be placed on the support of the object since it can be no larger than half the diameter of the autocorrelation along any direction. Additional measurements or other forms of a priori information may be available for specific applications, but in this paper we will restrict our attention to nonnegativity and support constraints in the object domain since they occur most frequently in real-world applications.

Until about a decade ago, there was much pessimism that the phase retrieval problem could be solved or that the solution would be useful, because the one-dimensional theory of analytic functions available at the time indicated that there were ordinarily an enormous number of ambiguous solutions [1-3].

The first indications that the 2-D case is usually unique, despite the lack of uniqueness in one dimension, came from empirical reconstruction results [4, 5]: images that were reconstructed resembled the original simulated objects used to compute the Fourier modulus data. Those results gave hope that 2-D phase retrieval problems might be solvable and unique. (Other phase retrieval problems, such as in electron microscopy where one has squared-modulus measurements in each of two domains [6] or in x-ray crystallography where one has the a priori information that the object consists of a finite collection of atoms [7], had been solved; but those earlier successes depended upon much greater object-domain constraints than just nonnegativity and support.) Those empirical results gave impetus to attempts to extend the 1-D theory to 2-D. Although progress has been made [8-13], the level of understanding of the 2-D problem has not yet matched that of the 1-D problem.

One of the most enlightening developments has been the work of Bruck and Sodin [14], which modeled the object distribution as an array of delta functions on a regular grid. The Fourier transform

can then be expressed as a polynomial of two complex variables, and the presence of ambiguity in the phase retrieval problem is equivalent to the factorability of the polynomial. This explains the vast difference between the 1-D and the 2-D cases, because polynomials (of degree two or greater) of a single complex variable are always factorable whereas polynomials of two (or more) complex variables are rarely factorable [14-16]. Other interesting results have been obtained by exploiting this discrete model. Fiddy, Brames and Dainty [17, 18] described an object support that, by virtue of Eisenstein's irreducibility theorem, guarantees uniqueness. Brames [19] showed that any discrete object having a support whose convex hull has no parallel sides is unique among objects with the same convex hull; so if the convex hull of the support of such an object is known a priori, then it is unique. For these cases, there also exists a closed-form recursive reconstruction algorithm [20, 21].

These results for discrete objects are not the entire story, however, for two important reasons. First, if a given nonfactorable polynomial is near enough (in an integrated mean-squared difference sense, relative to the amount of noise present in the data) to a factorable polynomial, then the ambiguous solutions will be consistent with the noisy Fourier modulus data. It is not presently known how close an arbitrary polynomial is, on the average, to a factorable polynomial. Second, most real-world applications are for continuous objects, and the discrete (sampled) model used by Bruck and Sodin, although it has led to valuable insights, does not capture all the important features of the continuous real world. The uniqueness results for discrete objects do not necessarily have analogs in the continuous world. Two things that can be said of the continuous case are that objects known to satisfy the holography condition [22] are unique, and objects known to consist of separated parts are likely to be unique [23-25].

Whether discrete or continuous, it is easy to make up cases that are ambiguous. If $g(x, y)$ and $h(x, y)$ are two functions of finite support with Fourier transforms $G(u, v)$ and $H(u, v)$ respectively, then the convolutions

$$f_1(x, y) = g(x, y) * h(x, y) \tag{3}$$

and

$$f_2(x, y) = g(x, y) * h^*(-x, -y) \tag{4}$$

are different objects as long as neither g nor h is centrosymmetric, and they have Fourier transforms

$$F_1(u, v) = G(u, v)H(u, v) \tag{5}$$

and

$$F_2(u, v) = G(u, v)H^*(u, v) \tag{6}$$

which have the same modulus,

$$|F_1(u, v)| = |F_2(u, v)| = |G(u, v)| \ |H(u, v)| \ . \tag{7}$$

By this convolutional (products in the Fourier domain) method, it is possible to make up an uncountably infinite number of amibiguous cases even though the theory indicates that ambiguity is rare in 2-D. Thus, the fact that the probability of any given object being ambiguous (the Fourier transform being factorable) is zero is not necessarily comforting. Consider that it is also true that any randomly chosen real number has probability zero of being a rational number (almost all are irrational numbers). Yet any real number, even if irrational, can be approximated arbitrarily well by a rational number.

There is concern that if a given object is close enough to an ambiguous object, then the given object may be unique in theory but

ambiguous in practice. The existence of ambiguous objects close to a given object is likely to cause the existence of local minima in which iterative reconstruction algorithms will become trapped. Current theory has not adequately addressed this question even for the discrete model.

In this paper, partial answers to two of the uniqueness questions are given by means of computer simulations and reconstruction experiments. In Section 2, evidence is given that an interesting class of objects are in practice usually unique, and in Section 3 evidence is given that adding noise to the data does not alter that conclusion.

2. UNIQUENESS EXPERIMENTS

One approach to determining whether most objects of interest can, in practice, be uniquely reconstructed from their Fourier moduli is to computer simulate the Fourier moduli for a number of objects of interest and attempt to reconstruct them from their Fourier moduli using a phase retrieval algorithm. This approach is possible due to the existence of a practical phase retrieval algorithm, the iterative Fourier transform algorithm [4, 26-28].

The iterative transform algorithm uses the available data, an estimate (possibly noisy) of the modulus of the object's Fourier transform and the available object-domain constraints. For the astronomy problem, the constraint on the object is that it is nonnegative. In addition, from the Fourier modulus, one can compute the autocorrelation of the object from which one can place upper bounds on the support of the object (only in special cases can the support of the object be determined from the support of its autocorrelation) [29]. The iterative transform algorithm involves the repeated application of data and constraints in the Fourier and object domains. It is described in greatest detail in Reference 28 and has numerous applications in addition to phase retrieval [30].

If there are multiple ambiguous solutions, then the iterative transform algorithm is likely to find any one of them, depending on the random numbers used as the initial estimate. In one series of tests for a case that was made up to have two ambiguous solutions, out of ten reconstructions, each using a different random initial estimate, four converged to one of the two solutions and six converged to the other. In another series of tests for a case that was made up to have four ambiguous solutions, in each case that the algorithm converged, it converged only to one of the four solutions, and it converged to each of the four solutions at least once [31]. Hence, it is felt that if the iterative algorithm is run several times, each time using a different random initial estimate, and if it converges to the same solution each time, then it is highly likely that the solution is unique.

In an earlier study [32], the Fourier transforms of several different pictures of satellites were computed, the Fourier transforms were weighted by the MTF due to a circular aperture to include the effects of diffraction, and images were reconstructed from the moduli of the weighted Fourier transforms. The results of this set of experiments were very encouraging, but they were less conclusive than what was desired since the algorithm has a tendancy to stagnate in a certain mode. By stagnation, it is meant that the algorithm gets stuck on an output image that is not a solution, i.e., that does not agree simultaneously with all the data and constraints. Stagnation can be interpreted as being trapped in a local minimum of an error metric. The mode of stagnation most prevalent for these types of objects is that having an output image that resembles the original object but with a pattern of stripes superimposed [33, 34]. That this is a point of stagnation rather than (an ambiguous) solution is readily seen from the fact that the pattern of stripes extends (although with reduced contrast) beyond the known support of the object and therefore violates the object-domain support constraint.

This stagnation problem was a difficulty for several years, but more recently methods have been developed for overcoming it: the removal of the stripes by the voting method and by the patching method [35]. With these algorithm improvements, the experiments with the satellite images were redone, and the results are shown in Figure 1.

Eight digitized photographs of satellites were used. Each was downsampled to fit within an array of 64 by 64 pixels which was imbedded in a 128 by 128 array. Each were Fourier transformed (using the fast Fourier transform) and the complex Fourier transform was multiplied (weighted) by the autocorrelation function of a circular aperture of diameter 62 pixels to simulate the effects of diffraction for incoherent imaging through a circular aperture. Reference objects were computed by inverse Fourier transforming the weighted Fourier transform. These reference objects, which correspond to diffraction-limited images, are shown in Figures 1(a), (c), (e), (g), (i), (k), (m), and (o). The modulus of the weighted Fourier transform was used as the Fourier domain data, and a very loose support constraint consisting of a 64 x 64 square was employed along with a nonnegativity constraint. The following sequence of steps were performed for each object. The initial input to the algorithm was an array of uniformly distributed random numbers. Twenty error-reduction iterations, followed by several cycles of one hundred hybrid input-output iterations (with the feedback parameter $\beta = 0.7$) plus forty error-reduction iterations each, followed by forty more error-reduction iterations were used to produce an output image [4, 28]. Two more output images were produced in a similar fashion starting with different random initial inputs. In case there might be a problem of stagnation with stripes, the three output images were combined by the voting method [35] and starting from that result another sequence of iterations like that described above was performed to produce a fourth output image. For the same reason, the

first and second output images were combined by the patching method [35] and the first and third output images were combined similarly. These two results were then combined by the patching method. Starting from that result, another sequence of iterations like the one described above was performed to produce a fifth output image. The object-domain error metric E_o (defined below) was computed for each of the five output images described above, and the one with the lowest error metric, i.e., the one in best agreement with the data and constraints, was taken to be the solution. The lengthy procedure and large number of iterations described above would ordinarily be overkill for this phase retrieval problem, but it was used in order to ensure a high probability of success using a totally automated procedure.

The resulting images are shown in Figure 1 next to the respective reference objects. Table 1 shows the normalized root-mean-squared error (NRMSE) of each defined by

$$E_o = \left[\frac{\sum_{(x,y)\in\gamma} |g'(x, y)|^2}{\sum_{(x,y)} |g'(x, y)|^2} \right]^{1/2} \tag{8}$$

where γ is the set of points at which the output image $g'(x, y)$ violates the constraints (all points outside the 64 x 64 support and all points within the support at which g' is negative). As shown in the second column of Table 1, the reference objects have nonzero error because the simulated telescope MTF causes sidelobes of the points within the object to extend beyond the support of the object. Note that the impulse response is a $J_1^2(x)/x^2$ function. If a reference object is used as a starting estimate for the iterative algorithm, it converges to a solution that is more consistent with the support constraint, but otherwise looks almost identical to the reference object. The NRMS errors of such images after 21

iterations of the error reduction algorithm are listed in the third column of Table 1. One would expect this to approximate the lowest error metric that any reconstruction algorithm would produce.

Reconstructed images (b), (d), (n) and (p) are visually excellent. Images (b) and (d) even have lower NRMSE than their respective reference objects (even after the 21 error reduction iterations). Presumably, this is possible because many more iterations were used to make the reconstructed images consistent with the support constraint. Reconstructed images (h) and (l) are in good agreement with the data and constraints and agree with the reference objects in gross detail, but they are still plagued by residual stripes that were not fully removed by the stripe-reducing methods -- (h) has coarse vertical stripes and (l) has a checkerboard pattern of at least two sets of stripes. Output image (j) has a strong pattern of residual diagonal stripes and is only in moderately good agreement with the data and constraints: its NRMSE is 0.236×10^{-2} versus 0.179×10^{-2} that is possible for this case. Nevertheless, image (j) closely resembles the reference object aside from the superimposed stripes. Reconstructed image (f) is the most troublesome. Despite being in moderately good agreement with the data and constraints, it does not closely resemble the reference object. Of this group of images, it is the only one that has a clear uniqueness problem. The fact that the reference object, shown in Figure 1(e), is nearly centrosymmetric may contribute to the difficulty in uniquely reconstructing it.

In summary, all but one of the eight reconstructed images appear to be, practically speaking, uniquely reconstructed. The one case that clearly suffers from nonuniqueness is a nearly centro-symmetric object (e-f). Objects that appear to have separated parts (a-b, c-d, m-n, o-p) were the easiest to reconstruct.

3. EFFECTS OF NOISE

Another major concern is the effect of noise on the uniqueness of the solution. Ordinarily, there is no solution consistent with the data and constraints when noise is present in the data. The iterative algorithm is an attempt to find an output image that is as consistent as possible with the conflicting noisy data and constraints. Analysis has indicated that the uniqueness condition is stable in the presence of noise [36]. In this section, experimental results [33] are reviewed that support that viewpoint.

For this set of experiments, stellar speckle interferometry [37] was simulated, including the effects of atmospheric turbulence, photon noise, and averaging a finite number of speckle frames. Images were reconstructed from the simulated noisy Fourier modulus data using the iterative transform algorithm.

The object used is that shown in Figure 1(o). It is about 64 by 40 pixels in extent imbedded in a 128 by 128 array. It was convolved with 156 different point-spread functions each simulating a different realization of the turbulent atmosphere. Figures 2(a) through 2(d) show four examples of the resulting blurred images. The blurred images were then subjected to a Poisson noise process to simulate the effects of photon noise. A variety of noise levels were assumed for different experiments. The same four blurred images, each with three different levels of photon noise, are shown in the remainder of Figure 2. For each photon noise level, the 156 degraded images were processed by Labeyrie's method [37] using the Goodman-Belsher [38] modification to suppress a noise bias term to arrive at a noisy estimate of the Fourier modulus of the object. Four examples with different photon noise levels are shown in Figure 3. Further details on the data simulation are available in Reference 33.

Images reconstructed from the noisy Fourier modulus data are shown in Figure 4. Figure 4(a) shows the diffraction-limited object

for comparison. When twin images were reconstructed, they were inverted to allow for easier comparison. The stripe removal methods were not employed for this series of experiments. The image shown in Figure 4(e) was reconstructed from the data derived from the case of 305,000 photons per degraded image which was calculated to be the expected number of photons available for the simulated imaging scenario. Each successive 3rd image in Figure 4 arose from data using approxmately one tenth as many photons. Thus, Figure 4(k), which still has some useful image information, was obtained from only 1/100 the number of photons that could be expected in practice. That the noisiest case, Figure 4(m), resulted in an unrecognizable image is what one would expect, since, as can be seen from Figure 3(d), the Fourier modulus data was so noisy as to be unrecognizable as well.

The encouraging conclusion derived from these results is that, as the noise in the Fourier modulus estimate increases, the reconstructed image degrades in a gradual and predictable manner. That is, small amounts of noise do not, as was once mistakenly predicted [39], cause the uniqueness of the solution to suddenly go unstable and cause radically different images to be reconstructed. On the contrary, these results give experimental evidence to the stability predicted more recently [36].

4. CONCLUSIONS

By simulating Fourier modulus data and performing reconstruction experiments, it has been shown that (1) most images of satellites are uniquely related to their Fourier modulus and (2) the uniqueness is stable in the presence of noise. These results should instill confidence that the solution to the phase retrieval problem by the iterative Fourier transform algorithm is realistically unique and noise tolerant in a practical sense, making it useful for real-world applications.

ACKNOWLEDGEMENT

The authors acknowledge the assistance of G.B. Feldkamp who produced the results shown in Section 3. This work was supported by the U.S. Air Force Office of Scientific Research under Contract F49620-82-K-0018 and in part by Rome Air Development Center under Contract F30602-80-C-0002.

REFERENCES

1. E. Wolf, "Is a Complete Determination of the Energy Spectrum of Light Possible from Measurements of the Degree of Coherence," Proc. Phys. Soc. (London) 80, 1269-1272 (1962).

2. A. Walther, "The Question of Phase Retrieval in Optics," Optica Acta 10, 41-49 (1963).

3. E.M. Hofstetter, "Construction of Time-Limited Functions with Specified Autocorrelation Functions," IEEE Trans. Info. Theory IT-10, 119-126 (1964).

4. J.R. Fienup, "Reconstruction of an Object from the Modulus of Its Fourier Transform," Opt. Lett. 3, 27-29 (1978).

5. P.J. Napier and R.H.T. Bates, "Inferring Phase Information from Modulus Information in Two-Dimensional Aperture Synthesis," Astron. Astrophys. Suppl. 15, 427-430 (1974).

6. W.O. Saxton, Computer Techniques for Image Processing in Electron Microscopy (Academic Press, New York, 1978).

7. G.H. Stout and L.H. Jensen, X-Ray Structure Determination (Macmillan, London, 1968).

8. W. Lawton, "A Numerical Algorithm for 2-D Wavefront Reconstruction from Intensity Measurements in a Single Plane," in 1980 International Optial Computing Conference, W. T. Rhodes, ed., Proc. SPIE 231, 94-98 (1980).

9. M. Nieto-Vesperinas, "Dispersion Relations in Two Dimensions: Application to the Phase Problem," Optik (Stuttgart) 56, 377-384 (1980).

10. I. Manolitsakis, "Two-Dimensional Scattered Fields: A Description in Terms of the Zeros of Entire Functions," J. Math. Physics 23, 2291-2298 (1982).

REFERENCES (Continued)

11. R. Barakat and G. Newsam, "Necessary Conditions for a Unique Solution to Two-Dimensional Phase Recovery," J. Math. Phys. 25, 3190-3193 (1984).

12. J.L.C. Sanz and T.S. Huang, "Unique Reconstruction of a Band-Limited Multidimensional Signal from its Phase or Magnitude," J. Opt. Soc. Am. 73, 1446-1450 (1983).

13. I.S. Stefanescu, "On the Phase Retrieval Problem in Two Dimensions," J. Math. Phys. 26, 2141-2160 (1985).

14. Yu.M. Bruck and L.G. Sodin, "On the Ambiguity of the Image Reconstruction Problem," Opt. Commun. 30, 304-308 (1979).

15. L. Carlitz, "The Distribution of Irreducible Polynomials in Several Indeterminates," Illinois J. of Math., 371-375 (1963).

16. M.H. Hayes and J.H. McClellan, "Reducible Polynomials in More Than One Variable," Proc. IEEE 70, 197-198 (1982).

17. M.A. Fiddy, B.J. Brames, and J.C. Dainty, "Enforcing Irreducibility for Phase Retrieval in Two Dimensions," Opt. Lett. 8, 96-98 (1983).

18. M. Nieto-Vesperinas and J.C. Dainty, "A Note on Eisenstein's Irreducibility Criterion for Two-Dimensional Sampled Objects," Opt. Commun. 54, 333-334 (1985).

19. B.J. Brames, "Unique Phase Retrieval with Explicit Support Information," Opt. Lett. 11, 61-63 (1986).

20. J.R. Fienup, "Reconstruction of Objects Having Latent Reference Points," J. Opt. Soc. Am. 73, 1421-1426 (1983).

21. T.R. Crimmins, "Phase Retrieval for Discrete Functions with Support Constraints," in Signal Recovery and Synthesis II, digest of papers for OSA topical meeting, Honolulu, Hawaii, 2-4 April 1986, pp. 75-78.

REFERENCES (Continued)

22. E.N. Leith and J. Upatnieks, "Reconstructed Wavefronts and Communication Theory," J. Opt. Soc. Am. 52, 1123-1130 (1962).

23. A.H. Greenaway, "Proposal for Phase Recovery from a Single Intensity Distribution," Opt. Lett. 1, 10-12 (1977).

24. T.R. Crimmins and J.R. Fienup, "Ambiguity of Phase Retrieval for Functions with Disconnected Support," J. Opt. Soc. Am. 71, 1026-1028 (1981).

25. T.R. Crimmins and J.R. Fienup, "Uniqueness of Phase Retrieval for Functions with Sufficiently Disconnected Support," J. Opt. Soc. Am. 73, 218-221 (1983).

26. J.R. Fienup, "Space Object Imaging Through the Turbulent Atmosphere," Opt. Eng. 18, 529-534 (1979).

27. J.R. Fienup, "Iterative Method Applied to Image Reconstruction and to Computer-Generated Holograms," Opt. Eng. 19, 297-305 (1980).

28. J.R. Fienup, "Phase Retrieval Algorithms: A Comparison," Appl. Opt. 21, 2758-2769 (1982).

29. J.R. Fienup, T.R. Crimmins, and W. Holsztynski, "Reconstruction of the Support of an Object from the Support of Its Autocorrelation," J. Opt. Soc. Am. 72, 610-624 (1982).

30. J.R. Fienup, "Reconstruction and Synthesis Applications of an Iterative Algorithm," in Transformations in Optical Signal Processing, W.T. Rhodes, J.R. Fienup, and B.E.A. Saleh, eds., Proc. SPIE 373, 147-160 (1981).

31. P. vanToorn, A.H. Greenaway and A.M.J. Huiser, "Phaseless Object Reconstruction," Optica Acta 7, 767-774 (1984).

32. J.R. Fienup, "Fourier Modulus Image Construction," Report RADCTR-81-63 (1981).

33. G.B. Feldkamp and J.R. Fienup, "Noise Properties of Images Reconstructed from Fourier Modulus," in 1980 International Optical Computing Conference, W.T. Rhodes, ed., Proc. SPIE 231, 84-93 (1980).

34. J.R. Fienup, "Experimental Evidence of the Uniqueness of Phase Retrieval from Intensity Data," in Indirect Imaging, Proceedings of URSI/IAU Symposium, 30 Aug. to 2 Sept. 1983, Sydney, Australia, ed. J.A. Roberts (Cambridge University Press, Cambridge, 1984), pp. 99-109.

35. J.R. Fienup and C.C. Wackerman, "Phase Retrieval Stagnation Problems and Solutions," submitted to J. Opt. Soc. Am. A.

36. J.L.C. Sanz, T.S. Huang, and F. Cukierman, "Stability of Unique Fourier-Transform Phase Reconstruction," J. Opt. Soc. Am. 73, 1442-1445 (1983).

37. A. Labeyrie, "Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images," Astron. and Astrophys. 6, 85-87 (1970).

38. J.W. Goodman and J.F. Belsher, "Fundamental Limitations in Linearly Invariant Restoration of Atmospherically Degraded Images," in Imaging Through the Atmosphere, Proc. SPIE 75, 141-154 (1976).

39. A.M.J. Huiser and P. Van Toorn, "Ambiguity of the Phase-Reconstruction Problem," Opt. Lett. 5, 499-501 (1980).

TABLE 1

NORMALIZED RMS ERROR METRIC OF REFERENCE OBJECTS AND IMAGES RECONSTRUCTED FROM FOURIER MODULUS USING THE ITERATIVE ALGORITHM. Letters in parentheses refer to Figure 1. All errors are in percents (X0.01).

Reference Object			Reconstructed Image	
Figure 1	NRMSE(1)	NRMSE(21)	Figure 1	NRMSE
(a)	0.252	0.184	(b)	0.164
(c)	0.468	0.359	(d)	0.310
(e)	0.215	0.194	(f)	0.223
(g)	0.316	0.240	(h)	0.240
(i)	0.210	0.179	(j)	0.236
(k)	0.319	0.207	(l)	0.233
(m)	0.377	0.275	(n)	0.349
(o)	0.386	0.328	(p)	0.576

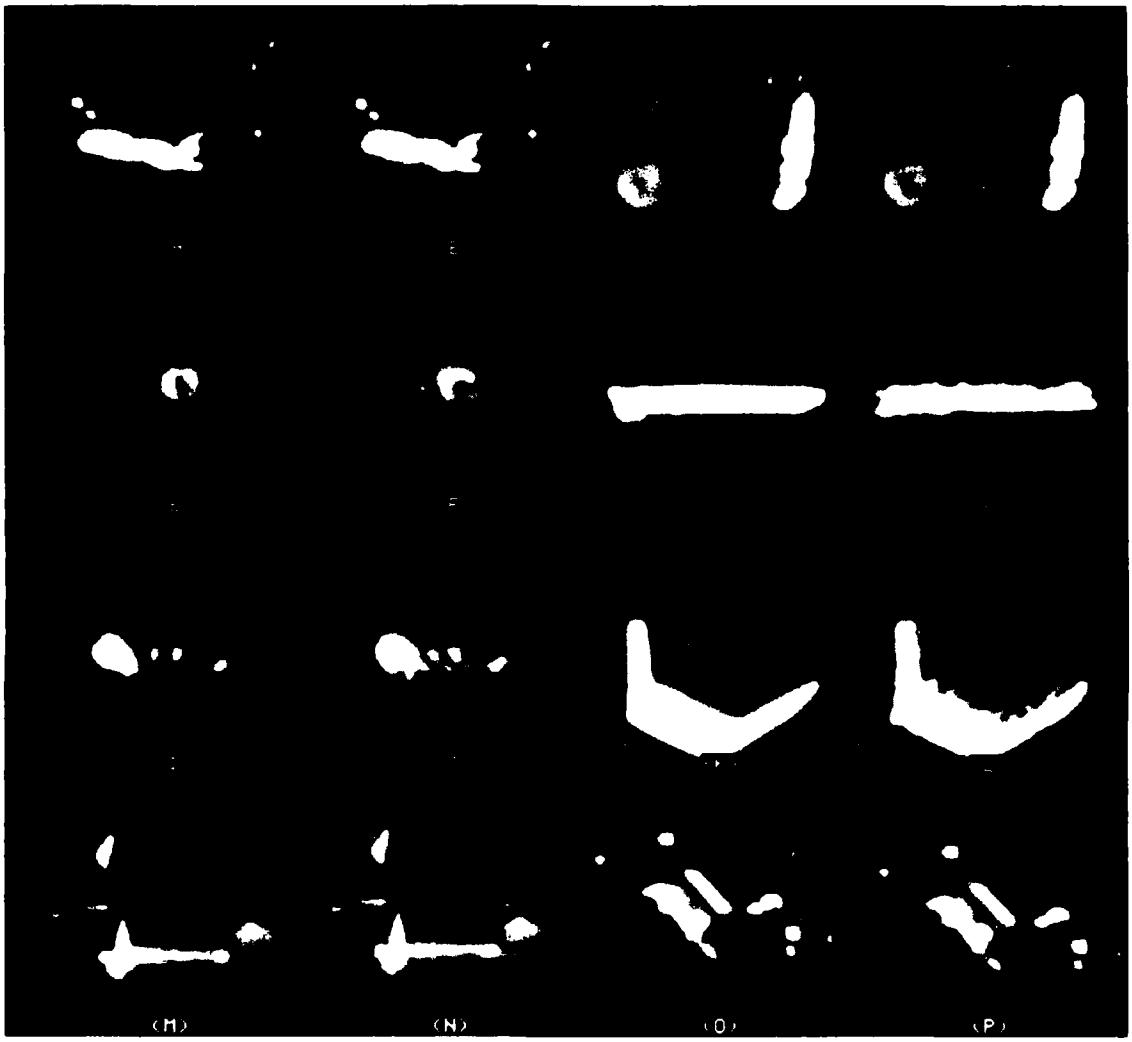


Figure 1. Diffraction-limited reference objects: (a), (c), (e), (g), (i), (k), (m), (o). Respective images reconstructed from the moduli of the Fourier transforms of the reference objects using the iterative Fourier transform algorithm: (b), (d), (f), (h), (j), (l), (n), (p).

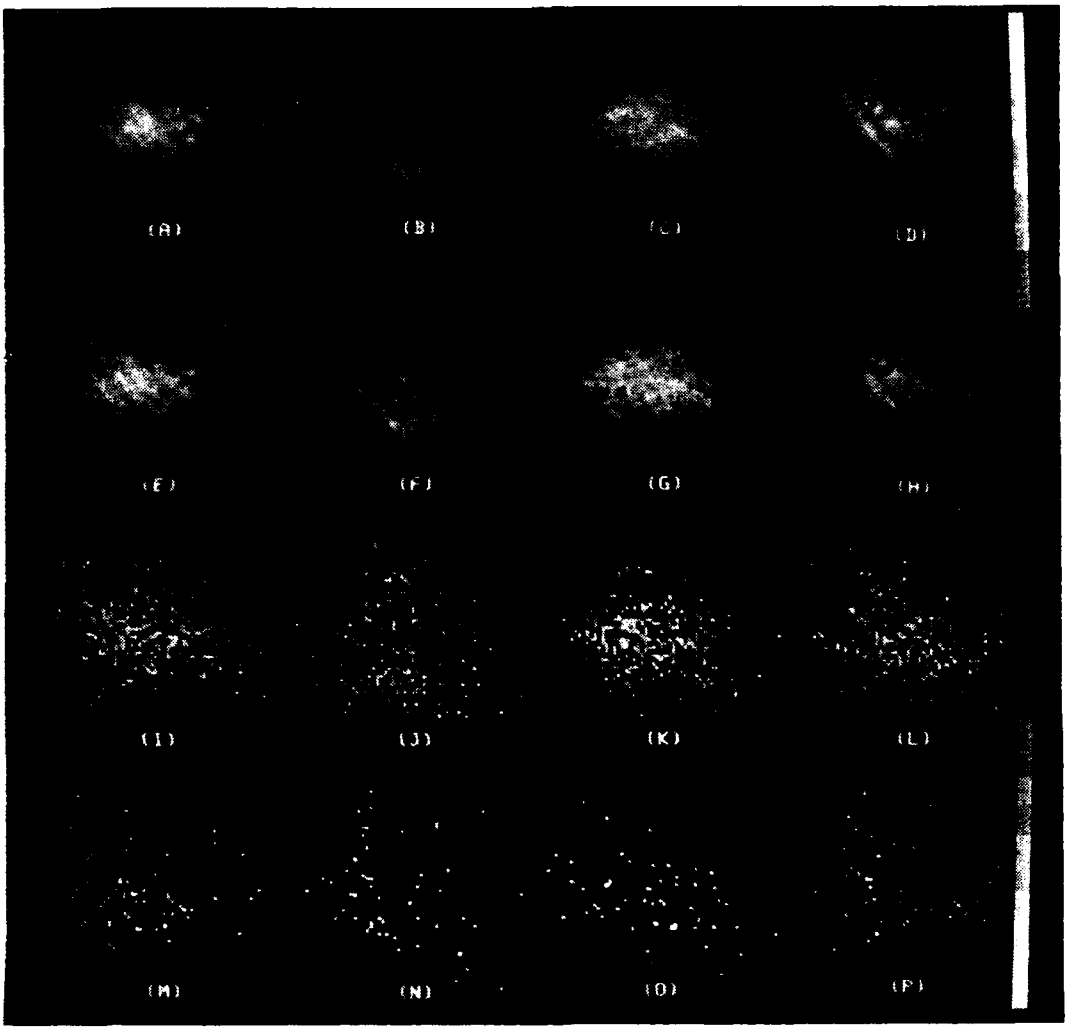


Figure 2. Images of the object shown in Figure 1(o) as would be seen through an optical telescope. (a) - (d) Four images, each blurred by a different realization of the turbulent atmosphere; blurred images including photon (Poisson) noise -- number of photons per degraded images: (e) - (h) 305,000; (i) - (l) 6143; (m) - (p) 643.

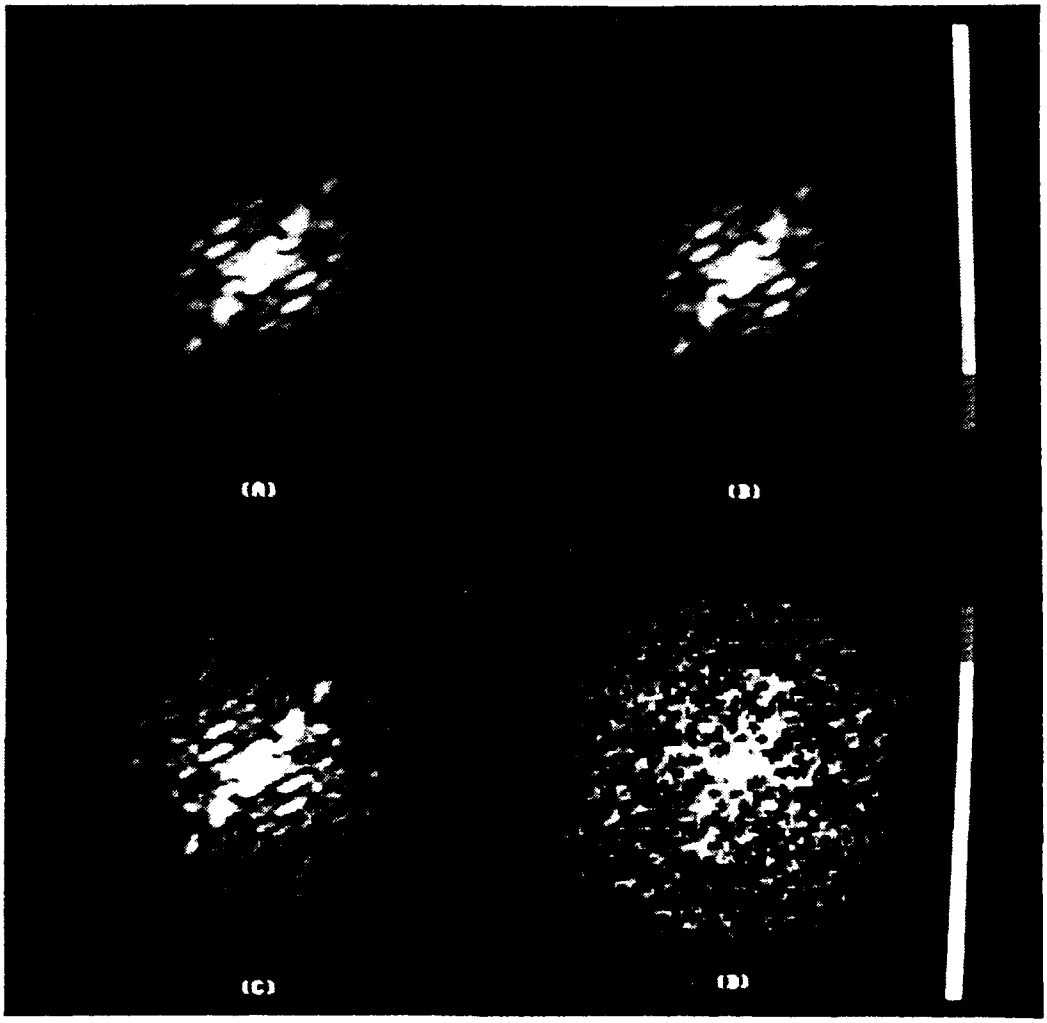


Figure 3. Fourier modulus estimates computed from 156 noisy, blurred images each. Number of photons per degraded image: (d) infinity (noise-free), (b) 305,000, (c) 6143, (d) 643.

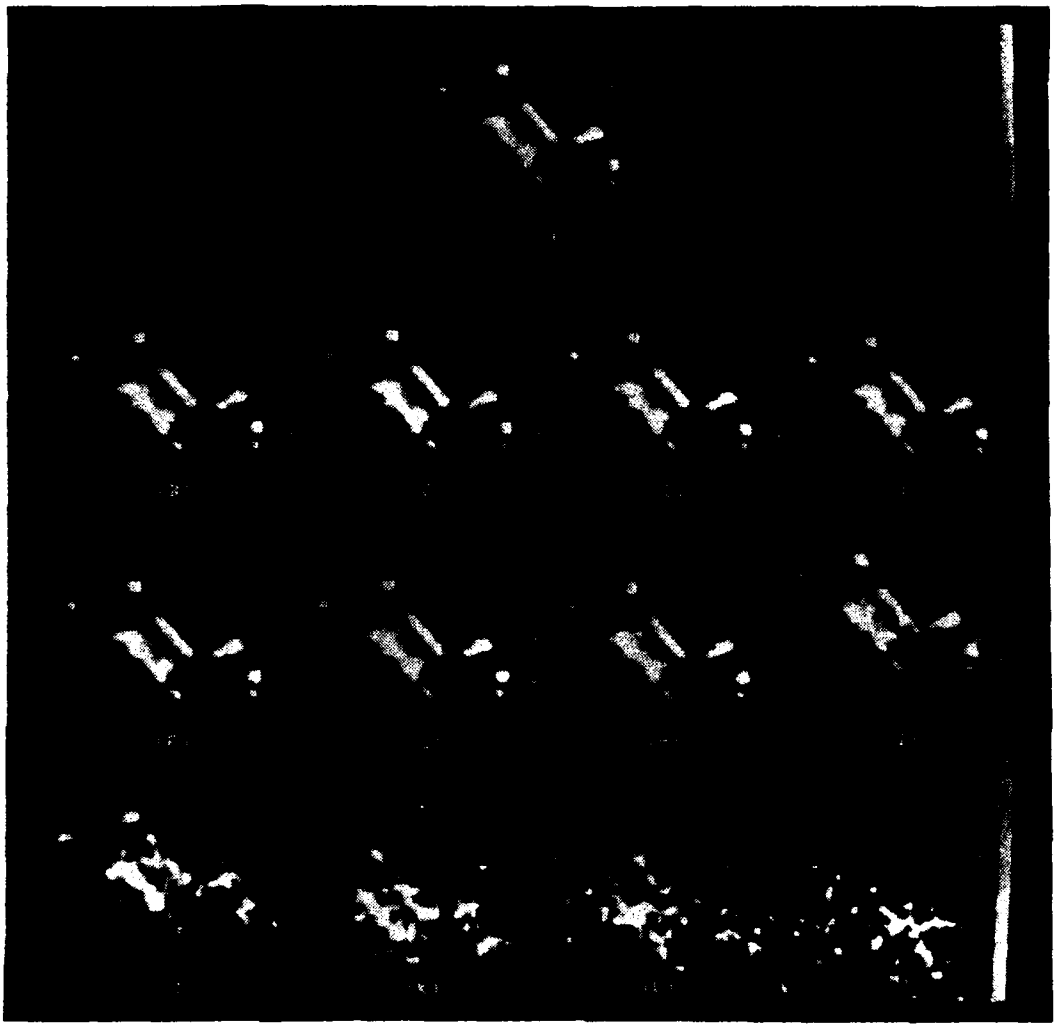


Figure 4. (a) Reference object; (b) - (m) images reconstructed from increasingly noisy Fourier modulus estimates using the iterative Fourier transform algorithm. Image (e) represents the case of 305,000 photons per degraded image (the realistic case), (j) 6143 photons, and (m) 643 photons.

Appendix E

RECONSTRUCTION OF A COMPLEX-VALUED OBJECT FROM THE MODULUS OF ITS FOURIER TRANSFORM USING A SUPPORT CONSTRAINT

J.R. Fienup

To appear in the Journal of the Optical Society of America A (January 1987).

Appendix E

Reconstruction of a Complex-Valued Object from the Modulus of Its Fourier Transform Using a Support Constraint

J. R. Fienup

Environmental Research Institute of Michigan

P.O. Box 8618, Ann Arbor, MI 48107

ABSTRACT

It is shown that, using the iterative Fourier transform algorithm, it is possible to reconstruct a complex-valued object from the modulus of its Fourier transform (i.e., solve the phase retrieval problem) using a sufficiently strong support constraint. Sufficiently strong support constraints include certain special shapes and separated supports. Reconstruction results are shown including the effect of tapered edges on the object's support.

To appear in the Journal of the Optical Society of America A (January 1987).

1. INTRODUCTION

In a number of disciplines, including astronomy, x-ray crystallography, electron microscopy and wavefront sensing, one encounters the phase-retrieval problem. One wishes to reconstruct f(x,y), an object function, from |F(u,v)|, the modulus of its Fourier transform

$$F(u,v) = |F(u,v)| \exp [i\psi(u,v)] = \mathcal{F}[f(x,y)] \qquad (1)$$

where $\mathcal{F}$ denotes Fourier transform. Since the autocorrelation of the object can be computed from the Fourier modulus by $\mathcal{F}^{-1}[|F(u,v)|^2]$, this problem is equivalent to reconstructing an object from its autocorrelation.

For successful reconstruction to be possible, one must have sufficiently strong a priori information (constraints) about the object to make the solution unique. Of course, one has the omnipresent ambiguities that f(x,y), $\exp(i\theta_c)f(x-x_0, y-y_0)$, and $\exp(i\theta_c)f^*(-x-x_0, -y-y_0)$, where θ_c is a constant phase, all have the same Fourier modulus. If these omnipresent ambiguities (phase constant, translation, and conjugate image) are the only ambiguities, then we consider the phase retrieval problem to be unique, and do not refer to these ambiguities as ambiguities. For most problems of interest (x-ray crystallography being a notable exception) the object function has finite support. (The support is the set of points over which the object is nonzero.) For the

case of 2-D sampled objects of finite support the solution is almost always unique (when no noise is present).[1,2] Also for 2-D continuous objects the solution is probably almost always unique, but the situation is less clear.[3,4]

The support of the object appears to play the most important role in determining whether the solution is unique. In 1-D the solution is almost always unique if the object is known to have support consisting of at least two disjoint parts satisfying a separation condition,[5] despite the fact that th 1-D case is usually not unique.[6] Therefore, one would expect the likelihood of ambiguity in the 2-D case to be lower for objects having supports known to have sufficiently separated parts. For some special supports the solution can be shown to be absolutely unique as opposed to "almost always unique." Objects including reference points known to satisfy the holography condition are unique. Sampled objects consisting of a rectangular region of support plus a point off one corner of the rectangle (for which one neighboring corner is non-zero) can be shown to be (absolutely) unique.[7-9] In addition, sampled objects known to have triangular support (with non-zero corners) and some other shapes including latent reference points are unique.[9] Recently this has been generalized to show that all sampled objects having supports whose known convex hulls are polygons having no parallel sides are unique[10,11] Another class of unique objects are those consisting of collections of delta functions with separations not satisfying certain redundancy conditions.[12] The library of supports for

which the solution is known to be unique is growing as we learn more about this important constraint.

Successful reconstruction requires, in addition to a likelihood of uniqueness, a phase retrieval algorithm that is not overly sensitive to noise and that converges to a solution using a reasonable amount of computer time. The most widely used phase retrieval algorithm satisfying these requirements is the iterative Fourier transform algorithm.[13-16] A descendent of the Gerchberg-Saxton type of algorithm,[17-19] it involves the transformation back and forth between the Fourier domain, where the Fourier modulus data is applied, and the object domain, where the a priori object constraints are applied.

For the astronomy problem the only a priori constraint is the object's nonnegativity. Since the autocorrelation of the object can be computed from $|F(u,v)|$, one also knows the support of the autocorrelation. For extended objects, from the support of the autocorrelation one can usually determine only upper bounds on the support of the object.[12] Therefore, the object domain constraints are nonnegativity and a loose support constraint. These constraints have been sufficient to reconstruct a number of computer-simulated astronomical-type objects,[13,14,20,21] even in the presence of a considerable amount of noise.[22,9,15]

In this paper we show that the iterative Fourier transform algorithm is also capable of reconstructing a complex-valued image from its Fourier modulus using only a support constraint if the support of

the object is sufficiently well known (is sufficiently tight), is sufficiently sharp (the object's edges are not tapered too much) and is one of several interesting types.

2. ITERATIVE FOURIER TRANSFORM ALGORITHM

For the problem under consideration in this paper it is assumed that $|F(u,v)|$ has been measured and the support, S, of $f(x,y)$ is known. The object $f(x,y)$ may be real (nonnegative or bipolar) or may be complex valued.

The k^{th} iteration of the iterative Fourier transform algorithm applied to this problem consists of the following four steps. (1) An input image, $g_k(x,y)$, is Fourier transformed yielding $G_k(u,v) = |G_k(u,v)| \exp [i\phi_k(u,v)]$; (2) a new Fourier-domain function is formed using the known Fourier modulus $|F(u,v)|$ with the computed phase: $G_k'(u,v) = |F(u,v)| \exp [i\phi_k(u,v)]$; (3) $G_k'(u,v)$ is inverse Fourier transformed to yield $g_k'(x,y)$; (4) a new input is formed by

$$g_{k+1}(x,y) = \begin{cases} g_k'(x,y), & (x,y)\in S \\ g_k(x,y)-\beta g_k'(x,y), & (x,y)\notin S \end{cases} \qquad (2)$$

where β is a constant usually chosen to be anywhere between 0.5 and 1.0 (the performance of the algorithm is not highly sensitive to the choice of the feedback parameter β[15]). Step (4) above embodies the hybrid input-output version of the iterative algorithm, which has been most successful for other phase retrieval problems.[15] The error-reduction

version of the iterative algorithm (which most closely follows the Gerchberg-Saxton philosophy) replaces Step (4) with

$$g_{k+1}(x,y) = \begin{cases} g_k'(x,y), & (x,y)\in S \\ 0, & (x,y)\notin S. \end{cases} \tag{3}$$

Other versions of the iterative algorithm are applicable as well.[15]

Progress of the algorithm can be monitored by the object-domain error metric, a normalized root-mean-squared (nrms) error (the amount by which the output image violates the object-domain constraint):

$$E_{ok} = \left(\frac{\sum_{(x,y)\notin S} |g_k'(x,y)|^2}{\sum_{(x,y)} |g_k'(x,y)|^2} \right)^{1/2} . \tag{4}$$

When using the error-reduction algorithm it is also appropriate to look at the Fourier domain error metric

$$E_{Fk} = \left(\frac{\sum_{(u,v)} [|G_k(u,v)| - |F(u,v)|]^2}{\sum_{(u,v)} |F(u,v)|^2} \right)^{1/2} . \tag{5}$$

The error-reduction algorithm can be proven to converge in the weak sense that[15]

$$E_{F(k+1)} \leq E_{ok} \leq E_{Fk} \leq E_{o(k-1)}. \tag{6}$$

Nevertheless, the error-reduction algorithm is usually much less effective than the hybrid input-output algorithm for which there is no convergence proof.[15] The error-reduction algorithm can be used in conjunction with the hybrid input-output algorithm in order to get a better reading of the residual error during the iterations.[15] Furthermore, when the reconstruction problem is particularly "easy", then even the error-reduction algorithm can perform adequately. The error-reduction algorithm is the same as the projection-onto-sets algorithm[23,24] for this problem; however, the modulus constraint in the Fourier domain is nonconvex and so the error-reduction algorithm does not enjoy the strong convergence properties of projection onto convex sets.

Note that Step (3) of the algorithm can be expressed as

$$G_k'(u,v) = G_k(u,v)\,[|F(u,v)|/|G_k(u,v)|]. \tag{7}$$

This form can be used instead of the form given in Step 2 above, but it obviously has a problem wherever $G_k(u,v) = 0$. This problem has always

been easily solved either by using

$$G_k'(u,v) = G_k(u,v)\,[|F(u,v)|/(|G_k(u,v)| + \delta)], \qquad (8)$$

where δ is a very small positive constant, or by directly setting $G_k'(u,v) = |F(u,v)| \exp[i\phi_k(u,v)]$, where $\phi_k(u,v)$ is the phase of $G_k(u,v)$. This has never caused any difficulties, contrary to speculation.[25]

3. RESULTS OF COMPUTER EXPERIMENTS

For all the reconstruction results shown, the reconstruction algorithm employed was the iterative Fourier transform algorithm using only a support constraint in the object domain. The initial input to the algorithm was an array of complex random numbers filling the area of the known support. In each case, first twenty iterations of the error-reduction algorithm were performed, then several cycles of iterations were performed, where one cycle of iterations is K iterations (K = 20 or 40) of the hybrid input-output algorithm with feedback parameter $\beta = 0.7$ followed by 10 iterations of the error-reduction algorithm. Quoted values of E_0 are after error-reduction iterations or after the end of a complete cycle.

Figures 1 and 2 show the results of computer experiments demonstrating the reconstruction of a complex-valued object from the modulus of its Fourier transform using only a support constraint and exploring the importance of a support constraint having separated parts.

The objects for these experiments were generated from a 64 x 64 portion of a complex-valued SEASAT SAR image of an area of land. A binary mask (an array of ones and zeros) was formed to define the desired support constraint. For the first case the support constraint was a pair of ellipses separated vertically by a distance greater than the sum of the vertical widths of the ellipses. This separation condition corresponds to the 1-D separation that makes uniqueness likely.[5] If this separation condition holds, then, within the complete autocorrelation of the object, the cross-correlation of the two ellipses does not overlap the autocorrelations of the individual ellipses. The complex-valued object, the modulus of which is shown in Figure 1(a), was formed by multiplying the SEASAT SAR image by the binary mask. The modulus of its Fourier transform, shown in Figure 1(c), was computed from the object imbedded in a 128 x 128 array. This imbedding is done in order to avoid aliasing in the computation of $|F(u,v)|^2$. The modulus of the complex-valued image reconstructed by the iterative Fourier transform algorithm using the Fourier modulus and the support constraint is shown in Figure 1(b). The reconstructed image is essentially perfect, both in modulus and in phase (not shown), up to an additive constant phase. Figure 2, curve 1, shows the object domain nrms error E_o, given by Eq. (4), as a function of iteration number. The output image looked excellent (visually indistinguishable from the original object) after the second cycle of iterations (120 total iterations). By iteration 820, E_o bottomed out at 3×10^{-8}, presumably limited by roundoff error. In practice with

noise-free data one would ordinarily stop the iterations once E_0 dropped below, say, 0.001.

Figures 1(d), 1(e), and 1(f) show the moduli of the object, the reconstructed image, and the Fourier transform, respectively, for a second case, for which the support is two larger, more closely-spaced ellipses not satisfying the separation condition of Reference 5. Again the reconstructed image is excellent, but a larger number of iterations were required to achieve an excellent output image (220 iterations, although it was very good by 170 iterations) than for the first case. This can be seen from iterations 120 and 170 of Figure 2, curve 2. E_0 continued to decrease by about a factor of two every 100 iterations and was 6×10^{-6} by iteration 1020.

Figures 1(g), 1(h), and 1(i) show the moduli of the object, the partially reconstructed output image after 1020 iterations, and the Fourier transform, respectively, for a third case, for which the support is contiguous, formed by the union of two overlapping ellipses. As seen from E_0 shown in Figure 2, curve 3, this output image does not agree exactly with the constraints and is not a solution. (Note that for this case E_0 is lower than it was for iteration 120 of the first case, for which the reconstruction was excellent. From case to case E_0 is often poorly correlated with how close the output image is to the original object -- it is only a measure of closeness to agreement with the data and constraints.) The algorithm was converging very slowly and was stopped before a solution was found. The quality of the partially

reconstructed image is very poor, although it does have some of the features of the object. This type of object appears to be much more difficult to reconstruct than one with separated support.

Figures 1(j), 1(k) and 1(l) show the moduli of the object, the reconstructed image, and the Fourier transform, respectively, for a fourth case, for which the support is donut-shaped with an off-center hole. Although the support is contiguous, the number of iterations required for convergence, as shown in Figure 2, curve 4, is similar to that of the objects having separated support. The output image looked very good by iteration 120 and excellent by iteration 170. By iteration 1020, E_0 had decreased to 1.5×10^{-7}. It is interesting to note that along any 1-D cut through the center of the object the support does have two separated parts.

Not shown is a fifth case for which the support was a single ellipse--the larger of the two ellipses shown in Figure 1(d). Similar to the third case above, the reconstruction was unsuccessful after several hundred iterations. For this case E_0 is shown in Figure 2, curve 5.

The results of Figures 1 and 2 demonstrate that having a support constraint consisting of (at least) two separated parts makes the reconstruction of the object by the iterative Fourier transform algorithm much easier than when using only a simple connected support constraint.

The power of the separated support might arise from its holographic-like properties. If one of the separated parts is point-like and sufficiently separated from the other parts, then it acts like a holographic reference point and the object can be easily extracted from its autocorrelation.[26,27] Latent reference points not satisfying the holography condition may also be used for some special cases.[9] One thing that a holographic reference point does is to encode the phase of the Fourier transform in a fringe pattern. For example, the object

$$f(x,y) = A\delta(x-x_0,y) + f_1(x,y) \tag{9}$$

has Fourier transform

$$F(u,v) = A \exp(-i2\pi u x_0) + F_1(u,v) \tag{10}$$

where $F_1(u,v) = |F_1(u,v)| \exp[i\psi_1(u,v)] = \mathcal{F}[f_1(x,y)]$.

Its squared Fourier modulus is

$$\begin{aligned} |F(u,v)|^2 &= |A \exp(-i2\pi u x_0) + F_1(u,v)|^2 \\ &= |A|^2 + |F_1(u,v)|^2 + 2|A|\,|F_1(u,v)| \cos[2\pi u x_0 + \psi_1(u,v)]. \end{aligned} \tag{11}$$

For separation x_0 sufficiently large compared with the width of the object, one can see the cos [ ] fringe in $|F(u,v)|^2$, and the spatial modulation of that fringe by the phase $\psi_1(u,v)$ gives an indication of the phase. If x_0 satisfies the holography condition, then $f(x,y)$ is trivially obtained by spatially filtering the autocorrelation function, which is given by $\mathcal{F}^{-1}[|F(u,v)|^2]$. From this it is seen that the phase $\psi_1(u,v)$ is obtained by taking the phase of $|F(u,v)|^2$ filtered by a single-sided bandpass filter.

For the problem under consideration one does not have holography since neither of the separated parts $f_0(x,y)$ and $f_1(x,y)$ of

$$f(x,y) = f_0(x,y) + f_1(x,y)$$

is necessarily a delta-function, nor is the separation necessarily large enough to satisfy the holography condition. (Furthermore, latent reference points[9] do not necessarily exist.) Nevertheless, with sufficiently separated parts one still does see a fringe-like structure in $|F(u,v)|^2$. As one departs further from the holography condition, both in terms of separation and in the greater extent of the smaller of the two parts, the fringes degrade into a speckle pattern, as seen by comparing Figures 1(c), 1(f) and 1(i). With a departure from the holography condition, the ability to decipher the phase from the degraded fringes diminishes. From the results shown in Figure 1 it appears that the iterative Fourier transform algorithm performs

especially well when the Fourier modulus data has any of the fringe structure described above, even when the fringes are substantially degraded.

Note that the supports were chosen to be non-centrosymmetric (except in the fifth case). This was done in order to avoid a potential stagnation problem that can occur for centrosymmetric supports since in this case $f^*(-x-x_0, -y-y_0)$ is consistent with the support constraint as well as having the same Fourier modulus as $f(x,y)$. As the iterations progress, the partially reconstructed image $g'(x,y)$ may possess features of both $f(x,y)$ and $f^*(-x-x_0, -y-y_0)$. It may be unable to move away from one of those equally valid solutions toward the other, and may stagnate in this condition. We have developed methods for overcoming this problem,[16,21] but reconstruction remains easier for non-centrosymmetric supports.

Figure 3 shows the results of computer experiments demonstrating the importance of the sharpness (or tapering) of the edges of an object. Figure 3(a) shows an object having triangular support and nonzero values in its three corners. These conditions ensure that the object is unique among objects having that support, and under these conditions there is a closed-form recursive algorithm for reconstructing it.[9] For the present experiment, the image, shown in Figure 3(b), was reconstructed using the iterative Fourier transform algorithm (rather than the recursive algorithm). Nonnegativity was not used as a constraint although the object happens to be nonnegative. In this case the algorithm converged

rapidly to the solution.

Another example of the ability to reconstruct an object with this type of support constraint using the iterative Fourier transform algorithm is shown in Reference 28. There the object was a pure-phase wavefront transmitted through a triangular aperture.

Since the recursive reconstruction algorithm and the uniqueness proof require the three corners to be nonzero, we wanted to determine the importance of nonzero corners to the iterative Fourier transform reconstruction algorithm. The same experiment was performed for the object shown in Figure 3(c), which is identical to the object shown in Figure 3(a) but with the corners zeroed out. The support constraint used in the iterative algorithm was the same triangular support as for the case above. The reconstructed image, shown in Figure 3(d), is the correct solution, but convergence was slower in this case than for the case of the object having three bright corners. Therefore, the brightness of the corners has an effect on convergence, but is not crucial as far as the iterative algorithm is concerned (they are crucial to the success of the recursive algorithm).

The effect of the sharpness of the edges of the object was also investigated. A third object having tapered edges, shown in Figure 3(e), was formed by multiplying the object shown in Figure 3(c) by a tapering function along each of its three edges. Thus this third object has the same triangular support as the other two cases described above, but it has small values near the edges of the support. An attempt was

made to reconstruct the image from its Fourier modulus using the iterative Fourier transform algorithm employing the triangular support constraint. The output image resulting after several hundred iterations is shown in Figure 3(f). Although the image is easily recognizable, it has a noisy appearance. It does not represent a solution since it is not in perfect agreement with the data and constraints. The algorithm was stagnating and the iterations were halted before a solution was found. This example shows that, if one does not use a nonnegativity constraint, then the sharpness of the edges of the object is very important to the ability of the iterative algorithm to reconstruct an image using only a support constraint in the object domain.

4. CONCLUSIONS

Previously it had been shown that by the iterative Fourier transform algorithm one could reconstruct a nonnegative object from the modulus of its Fourier transform employing a loose support constraint. In general the reconstruction of complex-valued objects is considerably more difficult than for real-valued, nonnegative objects.[29] The results shown here demonstrate the ability to reconstruct complex-valued objects if one has a tight enough support constraint that is one of a number of special types of support constraints. These special types of support constraints include supports having separated parts and supports for which the object can be reconstructed by the recursive algorithm using latent reference points; this latter class of objects includes objects with supports whose convex hulls have no parallel sides. One would

expect to be able to find other supports as well for which the iterative Fourier transform algorithm performs successfully. Simple symmetric support constraints such as single ellipses (circles) or rectangles do not work well. The algorithm also works much better for objects having sharp edges than for objects having tapered edges. Further research is being performed to examine in more detail the effects of the shape of the support, the amount of edge tapering, and the presence of noise on the ability to reconstruct a complex-valued object from the modulus of its Fourier transform using only a support constraint. Portions of this work were reported in References 21, 30 and 31.

Acknowledgement

The author acknowledges helpful discussions with C. C. Wackerman. This research was supported in part by the Air Force Office of Scientific Research under Contract F49620-82-K-0018 and in part under ERIM internal R&D funding.

REFERENCES

1. Yu. M. Bruck and L. G. Sodin, "On the Ambiguity of the Image Reconstruction Problem," Opt. Commun. 30, 304-308 (1979).

2. J. L. C. Sanz, T. S. Huang and F. Cukierman, "Stability of Unique Fourier-transform Phase Reconstruction, "J. Opt. Soc. Am. 73, 1442-1445 (1983).

3. I. Manolitsakis, "Two-Dimensional Scattered Fields: A Description in Terms of the Zeros of Entire Functions," J. Math. Physics 23, 2291-2298 (1982).

4. I. S. Stefanescu, "On the Phase Retrieval Problem in Two Dimensions," J. Math. Phys. 26, 2141-2160 (1985).

5. T. R. Crimmins and J. R. Fienup, "Uniqueness of Phase Retrieval for Functions with Sufficiently Disconnected Support," J. Opt. Soc. Am. 73, 218-221 (1983).

6. A. Walther, "The Question of Phase Retrieval in Optics," Optica Acta 10, 41-49 (1963).

7. M. A. Fiddy, B. J. Brames, and J. C. Dainty, "Enforcing Irreducibility for Phase Retrieval in Two Dimensions," Opt. Lett. 8,

96-98 (1983).

8. M. Nieto-Vesperinas and J.C. Dainty, "A Note on Eisenstein's Irreducibility Criterion for Two-Dimensional Sampled Objects," Opt. Commun. 54, 333-334 (1985).

9. J. R. Fienup, "Reconstruction of Objects Having Latent Reference Points," J. Opt. Soc. Am. 73, 1421-1426 (1983).

10. B. J. Brames, "Unique Phase Retrieval with Explicit Support Information," Opt. Lett. 11, 61-63 (1986).

11. T. R. Crimmins, "Phase Retrieval for Discrete Functions with Support Constraints: Summary," in Signal Recovery and Synthesis II Digest of Papers for the OSA Topical Meeting, 2-4 April, 1986, Honolulu, Hawaii, pp. 75-78.

12. J. R. Fienup, T. R. Crimmins, and W. Holsztynski, "Reconstruction of the Support of an Object from the Support of Its Autocorrelation," J. Opt. Soc. Am. 72, 610-624 (1982).

13. J. R. Fienup, "Reconstruction of an Object from the Modulus of Its Fourier Transform," Opt. Lett. 3, 27-29 (1978).

14. J. R. Fienup, "Space Object Imaging Through the Turbulent Atrmosphere," Opt. Eng. 18, 529-534 (1979).

15. J. R. Fienup, "Phase Retrieval Algorithms: A Comparison," Appl. Opt. 21, 2758-2769 (1982).

16. J. R. Fienup and C. C. Wackerman, "Improved Phase Retrieval Algorithm," J. Opt. Soc. Am. A 1, 1320 (A) (1984).

17. R. W. Gerchberg and W. O. Saxton, "A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures," Optik 35, 237-246 (1972).

18. W. O. Saxton, Computer Techniques for Image Processing in Electron Microscopy (Academic Press, New York, 1978).

19. R. W. Gerchberg, "Super-Resolution through Error Energy Reduction," Optica Acta 21, 709-720 (1974).

20. J. R. Fienup, "Experimental Evidence of the Uniqueness of Phase Retrieval from Intensity Data," in Indirect Imaging, Proceedings of URSI/IAU Symposium, 30 Aug. to 2 Sept. 1983, Sydney, Australia, ed. J. A. Roberts (Cambridge University Press, Cambridge, 1984), pp. 99-109.

21. J. R. Fienup, "Phase Retrieval: Algorithm Improvements, Uniqueness, and Complex Objects," in Signal Recovery and Synthesis II, Digest of Papers for the OSA Topical Meeting, 2-4 April 1986, Honolulu, Hawaii, pp. 40-43.

22. G. B. Feldkamp and J. R. Fienup, "Noise Properties of Images Reconstructed from Fourier Modulus,: in 1980 International Optical Computing Conference, W. T. Rhodes, ed., Proc. SPIE 231, 84-93 (1980).

23. D. C. Youla, "Generalized Image Restoration by Method of Alternating Orthogonal Projections," IEEE Trans. Circuits and Systems CAS-25, 694-702 (1978).

24. A. Levi and H. Stark, "Image Restoration by the Method of Generalized Projections with Application to Restoration from Magnitude," J. Opt. Soc. Am. A 1, 932-943 (1984).

25. J. L. C. Sanz, T. S. Huang and T-F. Wu, "A Note on Iterative Fourier Transform Phase Reconstruction from Magnitude," IEEE Trans. ASSP-32 1251-1254 (1984).

26. E. N. Leith and J. Upatnieks "Reconstructed Wavefronts and Communication Theory," J. Opt. Soc. Am. 52, 1123-1130 (1962).

27. J. W. Goodman, "Analogy Between Holography and Interferometric Image Formation," J. Opt. Soc. Am. 60, 506-509 (1970).

28. J. N. Cederquist, S. R. Robinson, D. Kryskowski, J. R. Fienup, and C. C. Wackerman, "Cramer-Rao Lower Bound on Fourier Modulus Wavefront Sensor," post-deadline paper to Signal Recovery and Synthesis II, OSA topical meeting, Honolulu, Hawaii, 2-4 April 1986.

29. R. H. T. Bates and D. G. H. Tan, "Fourier Phase Retrieval When the Image Is Complex," in Inverse Optics II, Proc. SPIE 558-10 (1985).

30. J. R. Fienup, "Phase Retrieval From a Single Intensity Distribution," invited paper to ICO-13, Sapporo, Japan, August 1984, in Optics in Modern Science and Technology, pp. 606-609.

31. J. R. Fienup, "Phase Retrieval Using a Support Constraint," IEEE ASSP Workshop on Multidimensional Digital Signal Processing, Leesburg, VA, 28-30 October 1985.

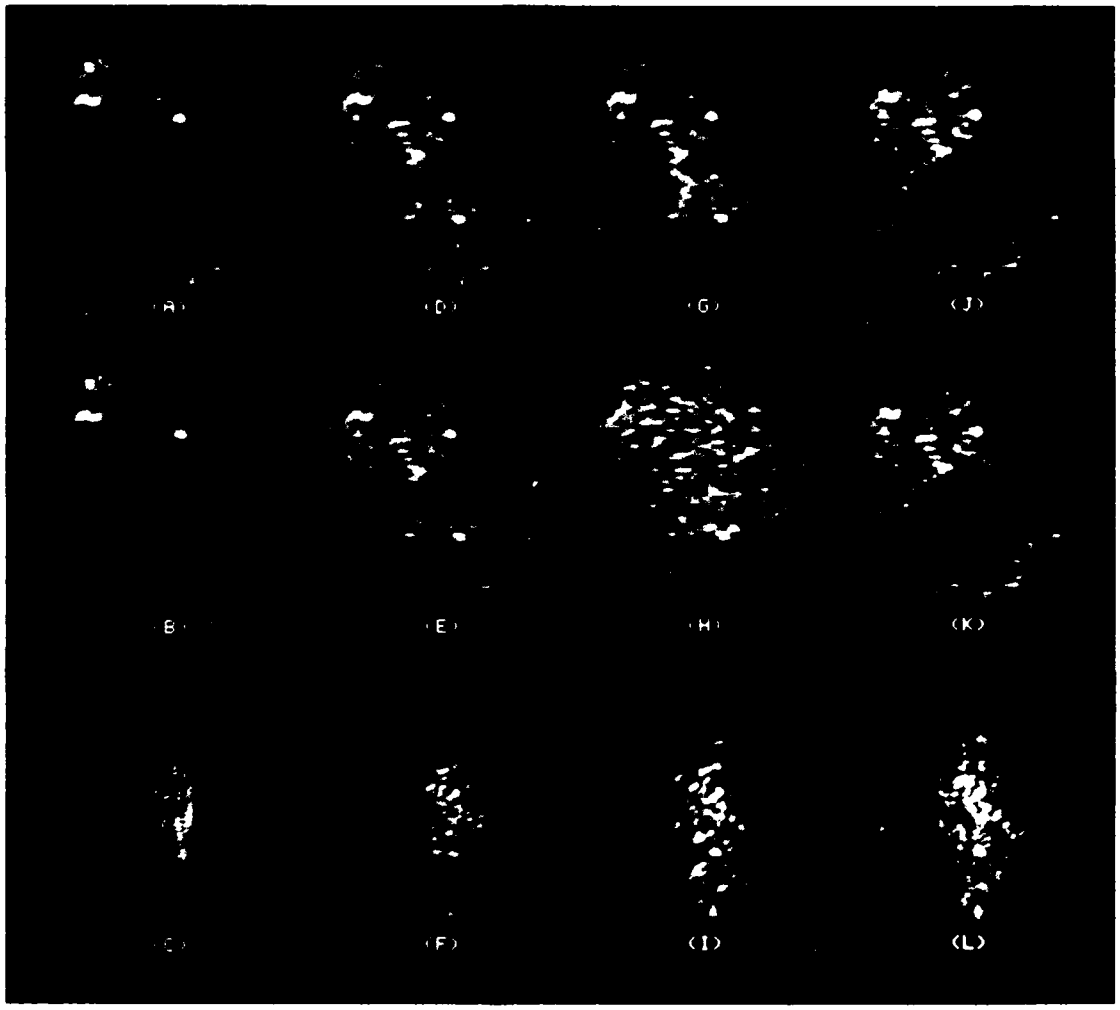


Figure 1. Examples of reconstructing complex-valued objects from the moduli of their Fourier transforms using a support constraint. (a), (d), (g), (j): moduli of the complex valued objects each having a different support; (b), (e), (h) and (k): the moduli of the images reconstructed using the iterative Fourier transform algorithm from the corresponding Fourier moduli, shown in (c), (f), (i) and (l).

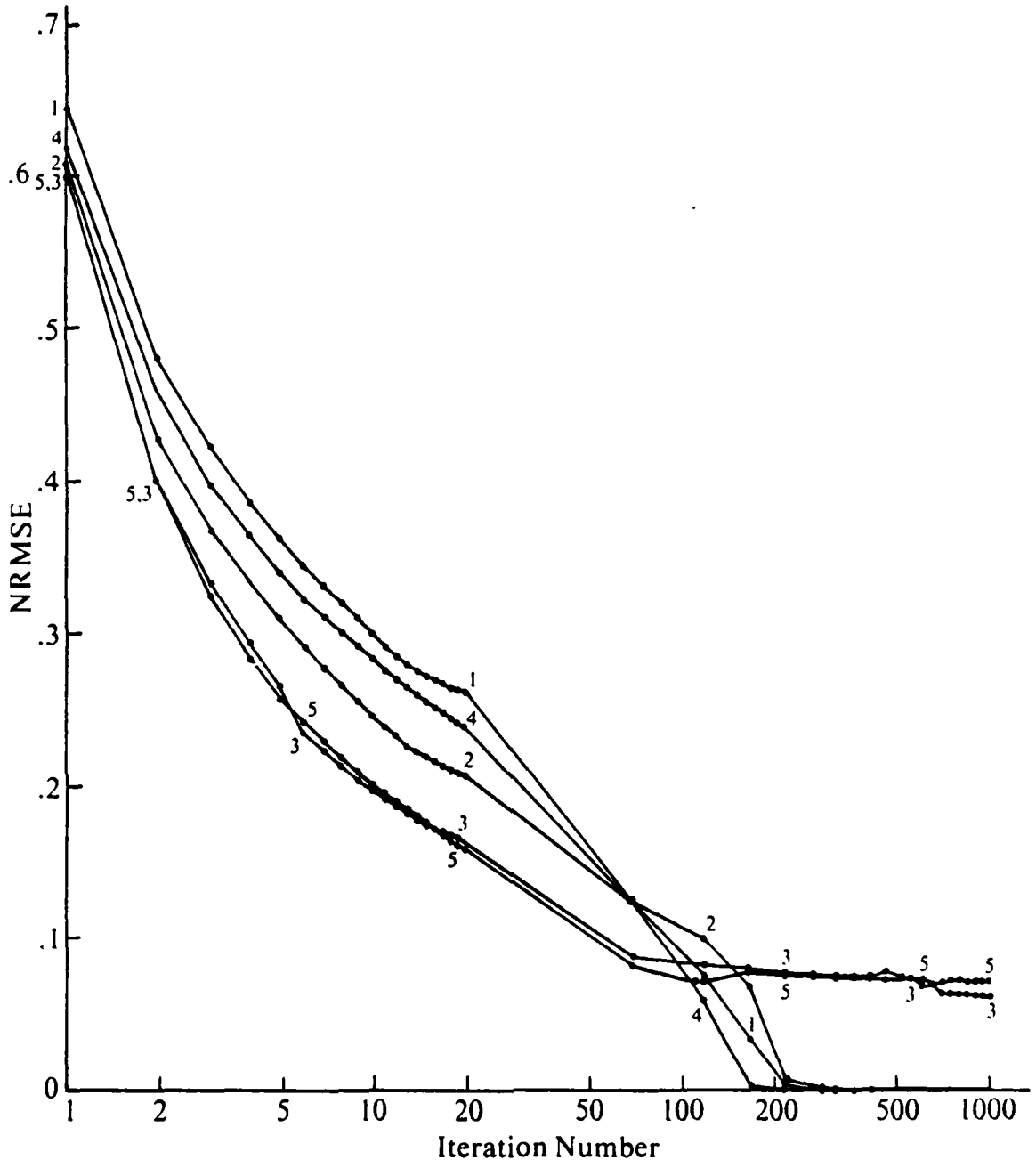
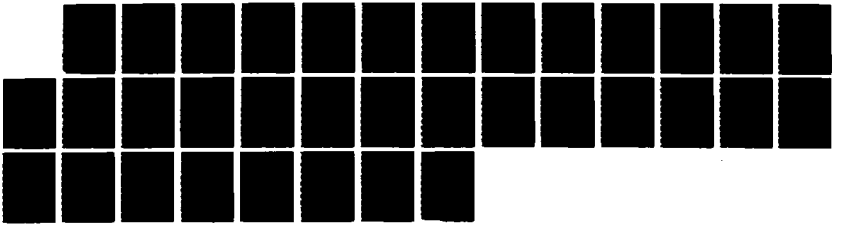
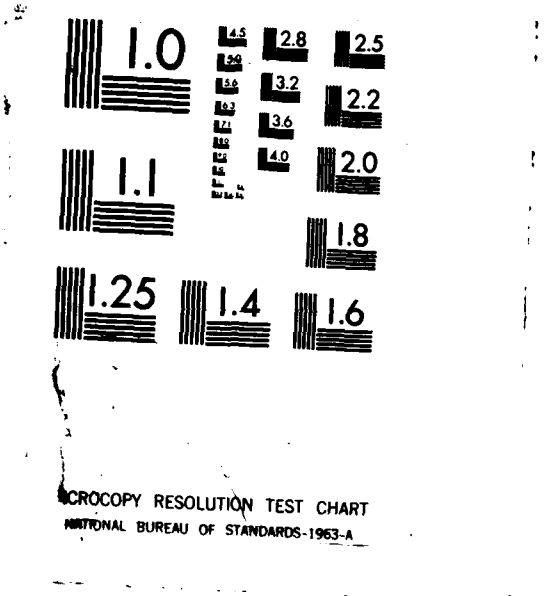


Figure 2. Error E_o as a function of iteration number for reconstruction examples. Curve 1 corresponds to the case of Figure 1(a), (b), (c); curve 2 to Figure 1(d), (e), (f); curve 4 to Figure 1(j), (k), (l); curve 5 to the case of the object being just the larger ellipse in Figure 1(d).

Figure 3. Examples of reconstructing objects from the moduli of their transforms using a triangular support constraint. (a) Object with sharp edges and bright corners, (b) reconstructed image; (c) object with sharp edges and zeroed corners, (d) reconstructed image; (e) object with tapered edges and zeroed corners, (f) partially reconstructed image.

AD-A174 100 DIFFRACTION-LIMITED IMAGING OF SPACE OBJECTS III(U) 3/3
ENVIRONMENTAL RESEARCH INST OF MICHIGAN ANN ARBOR
J R FIENUP ET AL OCT 86 ERIM-161900-20-F
UNCLASSIFIED AFOSR-TR-86-2109 F49620-82-K-0018 F/G 14/5 NL




1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
4.5
5.0
5.6
6.3
CROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

Appendix F

PHASE RETRIEVAL USING BOUNDARY CONDITIONS

J.R. Fienup

Reprinted from the Journal of the Optical Society of America A 3, 284-288 (February 1986).

Reprinted from Journal of the Optical Society of America A, Vol. 3, page 284, February 1986

Phase retrieval using boundary conditions

J. R. Fienup

Environmental Research Institute of Michigan, P. O. Box 8618, Ann Arbor, Michigan 48107

Received March 5, 1984; accepted September 19, 1985

It is shown that *a priori* knowledge of the edges of an object is not sufficient to ensure that it can be uniquely reconstructed from the modulus of its Fourier transform (or from its autocorrelation function). Furthermore, even in those cases for which the ultimate solution is unique, in intermediate steps in the solution by the recursive Hayes–Quatieri algorithm there can be ambiguities. An extension of the recursive algorithm that finds the solution (or solutions) is suggested, and it is shown that the recursive method can be applied to complex-valued objects.

INTRODUCTION

In a number of disciplines, including astronomy, x-ray crystallography, electron microscopy, and wave-front sensing, one encounters the phase-retrieval problem. One wishes to reconstruct $f(m, n)$, an object function, from $|F(p, q)|$, the modulus of its Fourier transform, where

$$F(p, q) = |F(p, q)|\exp[i\psi(p, q)] = \mathcal{F}[f(m, n)]$$
$$= \sum_{m=0}^{P-1}\sum_{n=0}^{Q-1} f(m, n)\exp[-i2\pi(mp/P + nq/Q)], \quad (1)$$

where $m, p = 0, 1, \ldots, P-1$ and $n, q = 0, 1, \ldots, Q-1$. The discrete transform is employed here since in practice one deals with sampled data in a computer. The problem of reconstructing the object from its Fourier modulus is equivalent to that of reconstructing the Fourier phase $\psi(p, q)$ from the Fourier modulus, since once one has the phase as well as the modulus, one can easily compute $f(m, n)$ by the inverse (discrete) Fourier transform (hence the name phase-retrieval problem). $r(m, n)$, the (aperiodic) autocorrelation of $f(m, n)$, is given by[1]

$$r(m, n) = \sum_{j=0}^{M-1}\sum_{k=0}^{N-1} f(j, k)f^*(j - m, k - n) \quad (2)$$

$$= \sum_{j=0}^{M-1}\sum_{k=0}^{N-1} f^*(j, k)f(j + m, k + n) \quad (3)$$

$$= \mathcal{F}^{-1}[|F(p, q)|^2], \quad (4)$$

where the asterisk denotes a complex conjugate and where it is assumed that $f(j, k) = 0$ for m outside $[0, M-1]$ and for n outside $[0, N-1]$. Note that, when simulating data, in order to avoid aliasing in the computation of $|F(p, q)|^2$ it is necessary that $M \le P/2$ and $N \le Q/2$. Since the autocorrelation function is easily computed from the Fourier modulus by Eq. (4), the phase-retrieval problem is equivalent to reconstructing an object from its autocorrelation function.

Several phase-retrieval algorithms have been demonstrated, all of them requiring some additional measurements or constraints on the solution. Examples include a reference point at least one object-diameter from the object[2] (giving rise to the holography condition[3]), a second intensity measurement in another plane[4,5] (in electron microscopy or wave-front sensing), nonnegativity and limited spatial extent[6,7] (in astronomy), just limited spatial extent,[8] atomic models[9] (in x-ray crystallography), objects consisting of collections of points having nonredundant spacings,[10] and objects having latent reference points[11] (not satisfying the holography condition). For some of these situations there is a proof of uniqueness of the solution that relies on the types of measurements made, on the *a priori* information available, or on the nature of the reconstruction algorithm itself.

Another proposed phase-retrieval algorithm is the Hayes–Quatieri (H-Q) recursive algorithm, which relies on *a priori* knowledge of the boundary conditions (i.e., the values of the edges of the object).[12,13] The purpose of this paper is to clarify the uniqueness questions pertaining to the H-Q recursive algorithm and to suggest a revised algorithm that finds the solution (or solutions) when the H-Q algorithm fails. The algorithm may also be applied to complex-valued objects. It is also pointed out that by the approach of using latent reference points,[11] special classes of objects can be shown to be unique for both real-valued and complex-valued objects.

AMBIGUITY USING BOUNDARY CONDITIONS

In Refs. 12 and 13 the H-Q recursive algorithm was put forward for reconstructing an object from the modulus of its Fourier transform, through the autocorrelation function, using boundary conditions, i.e., assuming knowledge of the edges of the object. A real-valued object, $f(m, n)$, was assumed to be zero outside the rectangular region of support $0 \le m \le M-1$ and $0 \le n \le N-1$. The top and bottom nonzero rows, $\beta(m) = f(m, N-1)$ and $\alpha(m) = f(m, 0)$, respectively, and the leftmost and rightmost nonzero col-

0740-3232/86/020284-05$02.00

umns, $f(0, n)$ and $f(M-1, n)$, respectively, are assumed to be known *a priori*. Rows 1 and $N-2$ can then be determined by solving a system of $2M-1$ linear equations in $2M-4$ unknowns. From Eq. (3) we have, for $n = N-2$, the second from the top row of the autocorrelation:

$$r(m, N-2) = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f^*(j, k) f(j+m, k+N-2)$$

$$= \sum_{j=0}^{M-1} f^*(j, 0) f(j+m, N-2)$$

$$+ \sum_{j=0}^{M-1} f^*(j, 1) f(j+m, N-1)$$

$$= \sum_{j=0}^{M-1} \alpha^*(j) f(j+m, N-2)$$

$$+ \sum_{j=0}^{M-1} f^*(j, 1) \beta(j+m) \qquad (5)$$

for $m = -M+1, \ldots, M-1$. These are $2M-1$ equations, one for each value of m, in $2M-4$ unknowns, $f(j, N-2)$ and $f(j, 1)$, for $j = 1, 2, \ldots, M-2$. Recall that $\alpha(j)$, $\beta(j)$, $f(0, N-2)$, $f(M-1, N-2)$, $f(0, 1)$, and $f(M-1, 1)$ are assumed to be known. After $f(j, N-2)$ and $f(j, 1)$ are determined by solving the system of equations given in Eqs. (5) above, then one can solve for $f(j, N-3)$ and $f(j, 2)$ using $r(m, N-3)$ in a similar manner. The remaining rows of the object are solved recursively in a similar manner.

The H-Q algorithm described above could work if the systems of equations had a unique solution for the unknowns. Restricting the solution to real-valued f's, a claim has been made that "It may be shown, however, that a sufficient condition for a unique solution . . . to exist is that $\alpha(m)$ and $\beta(m)$ not be identically zero and that $\alpha(m)$ not be related to $\beta(M-1-m)$ by a constant scale factor."[12] However, no proof of that statement was provided. In what follows, three examples that clarify this situation are given. In the first, the underlying phase-retrieval problem is not unique, yet the two ambiguous solutions have the same boundary values that satisfy the conditions quoted above. Therefore we have the unexpected result that, although for two-dimensional (2-D) sampled objects the phase-retrieval problem is usually unique to begin with,[14] knowledge of the boundary values is *not* sufficient to guarantee uniqueness. In the second example, the underlying phase-retrieval problem is unique, and the boundary values of the object satisfy the conditions quoted above, yet the equations to be solved for the H-Q recursive method do not yield a unique solution, contrary to the claim quoted above. An extension of the method that finds the solutions for the first two examples is given. In the third example, the H-Q algorithm is shown to work for a complex-valued object although the method was originally limited to real-valued objects.[12]

Example 1

Figures 1(a) and 1(b) show two different sampled objects having the same boundaries, and for both objects $\alpha(m)$ is not proportional to $\beta(M-1-m)$, and yet they have the same Fourier modulus and the same autocorrelation function, which is shown in Fig. 1(c). Therefore knowledge of the boundaries is not necessarily sufficient information for a unique reconstruction even if the restrictive condition quoted above is satisfied.

2	3	3	3	1
3	7	4	4	2
3	6	5	4	2
1	2	2	2	1

(a)

2	3	3	3	1
3	3	6	6	2
3	4	5	6	2
1	2	2	2	1

(b)

2	7	13	19	21	17	11	5	1
7	27	52	77	88	73	48	23	5
13	52	100	153	179	147	97	48	11
15	63	123	188	230	188	123	63	15
11	48	97	147	179	153	100	52	13
5	23	48	73	88	77	52	27	7
1	5	11	17	21	19	13	7	2

(c)

Fig. 1. Example 1. Two different objects, (a) and (b), have the same boundary values and also have the same Fourier modulus (not shown) and the same autocorrelation (c).

An infinite number of ambiguous examples such as that shown in Fig. 1 can be generated. From the theory of Bruck and Sodin[14] it is known that the solution of the phase-retrieval problem [but not necessarily of Eqs. (5)] is unique unless the Fourier transform of the object is a factorable polynomial, which is unlikely to happen by chance for the 2-D case. Factorability of the Fourier transform is equivalent to the object's being expressible as a convolution of two functions, and so ambiguous cases can be constructed by forming an object by convolving (or cross correlating) two functions.[15] The object in Fig. 1(a) was fabricated by cross correlating the functions shown in Figs. 2(a) and 2(b). The ambiguous solution shown in Fig. 1(b) is the inverted convolution of the functions shown in Figs. 2(a) and 2(b). An infinite number of other examples that are ambiguous even if one knows *a priori* the boundary values (and that satisfy the condition quoted above) can be obtained by replacing the values 1, 1, 1, and 2 of the function shown in Fig. 2(b) by other values w, x, y, and z, respectively, as long as $wz \neq xy$.

The H-Q recursive algorithm involves the solution of $2M-1$ linear equations in $2M-4$ unknowns.[12] One problem with this is that for $m = -M+1$ and for $m = M-1$, Eqs. (5) involve only the known boundary values and not the unknowns. Therefore one really has only $2M-3$ linear equations in $2M-4$ unknowns to begin with. A second problem is that on inspection of those equations one finds that, for the ambiguous cases such as that show in Fig. 1, two or more

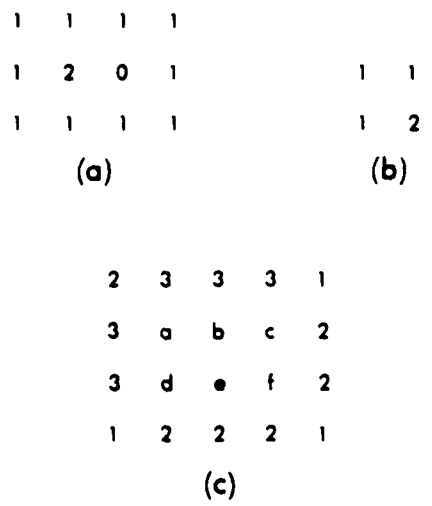


Fig. 2. Functions (a) and (b), which generate the object shown in Fig. 1(a) by cross correlation and in Fig. 1(b) by convolution. In (c) is the general form of the objects that have the autocorrelation shown in Fig. 1(c).

of them are dependent equations. Since the number of remaining linear independent equations is fewer than the number of unknowns, the problem is underdetermined, and multiple solutions exist.

Consider the particular example of Fig. 1(c), for which one searches for solutions of the form shown in Fig. 2(c), having the *a priori* known boundary values. The $2M - 3 = 7$ linear equations of Eq. (5), utilizing the second row of Fig. 1(c), are as follows (after rearranging the right-hand sides):

$$27 = 12 + a + 2f, \tag{6a}$$

$$52 = 12 + 2a + b + 2e + 3f, \tag{6b}$$

$$77 = 12 + 2a + 2b + c + 2d + 3e + 3f, \tag{6c}$$

$$88 = 13 + 2a + 2b + 2c + 3d + 3e + 3f, \tag{6d}$$

$$73 = 13 + a + 2b + 2c + 3d + 3e + f, \tag{6e}$$

$$48 = 13 + b + 2c + 3d + e, \tag{6f}$$

$$23 = 13 + c + d. \tag{6g}$$

Note that Eq. (6d) is equal to Eq. (6c) plus Eq. (6g), Eq. (6e) is equal to Eq. (6f) plus Eq. (6b) minus Eq. (6a), and Eq. (6c) is equal to Eq. (6f) plus Eq. (6b) minus Eq. (6g). That is, of these seven equations, three are dependent, leaving only four independent equations in six unknowns. Therefore one can, for example, choose values a and b in Fig. 2(c) arbitrarily, and then the values of c, d, e, and f are determined from Eqs. (6) as follows:

$$c = (15 - a + 2b)/4, \tag{7a}$$

$$d = (25 + a - 2b)/4, \tag{7b}$$

$$e = (35 - a - 2b)/4, \tag{7c}$$

$$f = (15 - a)/2. \tag{7d}$$

At this point the H-Q algorithm would have been stopped, leaving this ambiguity. An alternative is to continue by adding, to the existing set of underdetermined equations, the equations corresponding to the next row of the autocorrelation. Since the previous set of equations was underdetermined, some of this set of equations include terms that are the products of two unknowns, i.e., some of the equations are nonlinear. The first two such equations are (after rearranging the right-hand sides)

$$52 = 12 + 2a + 2c + d + 3f \tag{8a}$$

and

$$100 = 12 + 4b + 3c + 2d + 4e + af. \tag{8b}$$

Combining Eq. (8a) with Eqs. (7) yields

$$a = 15 - 2b, \tag{9}$$

and combining Eq. (8b) with Eqs. (7) and (9) yields

$$b^2 - 10b + 24 = 0 \tag{10a}$$

or

$$b = 4 \text{ or } 6. \tag{10b}$$

Evaluation of the other unknowns by Eqs. (9) and (7) gives the two solutions shown in Figs. 1(a) and 1(b). The equations for the unused points in the autocorrelation are found to be consistent with both solutions. Note that the steps described above for solving for the unknowns is a comprehensive procedure that finds all possible solutions. This modified approach can be generalized as follows. One performs the H-Q algorithm solving the sets of linear equations such as Eqs. (5), using the suggested pseudoinverse matrix[12] or another method such as Gauss elimination. If the number of independent equations is found to be exceeded by the number of unknowns, then one adds to the system of equations additional (possibly nonlinear) equations for the autocorrelation, using points in the next row of the autocorrelation. More equations are added (possibly using points unused by the H-Q algorithm) until the solution is unique or until all the points in the autocorrelation are used (in which case one might be left with multiple solutions). This modification to the H-Q algorithm may be difficult in some cases since it involves the solution of a system of nonlinear equations. Nevertheless, this approach is capable of finding all the solutions when multiple solutions exist. It works even when $\alpha(m)$ is proportional to $\beta(M - 1 - m)$.

Example 2

Figure 3(a) shows an object identical to that shown in Fig. 1(a) except that a value of 4 was replaced by a value of 5 in the fourth column from the left and second row from the top. Figure 3(b) shows its autocorrelation. As we show below, this object is uniquely related to its autocorrelation. On attempting to reconstruct the object from its autocorrelation by the H-Q recursive method using *a priori* knowledge of the boundary values, one finds, similar to the case of Example 1, using Gauss elimination, that three of the seven linear equations of Eqs. (5) are dependent, leaving only four independent equations in six unknowns. Again one can, for example, choose values a and b in Fig. 2(c) arbitrarily, and then the values of c, d, e, and f are determined. At this point the H-Q algorithm would have been stopped, leaving this ambiguity. Similar to the case of Example 1, an alternative is to

2	3	3	3	1
3	7	4	5	2
3	6	5	4	2
1	2	2	2	1

(a)

2	7	13	19	21	17	11	5	1
7	27	52	78	90	75	50	24	5
13	54	103	158	186	153	103	51	11
15	66	130	194	239	194	130	66	15
11	51	103	153	186	158	103	54	13
5	24	50	75	90	78	52	27	7
1	5	11	17	21	19	13	7	2

(b)

Fig. 3. Example 2. An object (a), which is uniquely related to its autocorrelation function (b). For this example Eqs. (5) do not have a unique solution.

carry on and continue to solve for the six unknowns using the equations for the next row of autocorrelation. Then one arrives at only one consistent solution, equal to the original object shown in Fig. 3(a).

Since the modified procedure described above finds all possible solutions having the given boundary values, and since only a single solution was found, the object shown in Fig. 3(a) is the unique solution.

An infinite number of such examples can be generated. If, instead of 5, any other value (except 4) had been used for the fourth column, second row of the object shown in Fig. 3(a), then the same behavior as in Example 2 would be observed: although the underlying phase-retrieval problem is unique, the solution to Eqs. (5) is not.

From the examples discussed above it is seen that the H-Q recursive algorithm, when modified to carry on with unknown variables as suggested above, is reminiscent of the recursive algorithm of Dallas,[5] except that Dallas had a tree of discrete solutions that grew with each recursive step, and ambiguities were resolved when the tree was pruned in later recursive steps. In the present case one must go deeper into the autocorrelation function than suggested in Ref. 12 to obtain enough independent equations to arrive at a solution (or solutions).

Example 3

In the preceding examples and in Refs. 11 and 12 it was assumed that the object was real valued. However, the reconstruction method can be just as well applied to complex-valued objects. Equations (5) simply becomes a system of $2M - 3$ linear equations having complex coefficients in $2M - 4$ complex-valued unknowns. (Alternatively, expressing the real and imaginary parts of each complex-valued equation as two separate equations and solving for the real and imaginary parts of the unknowns, one has a system of $4M - 6$ real linear equations in $4M - 8$ real unknowns. Dealing with the fewer number of equations with complex values is the simpler method.) In either case, if too many of the equations are linearly dependent, then one must resort to using additional (possibly nonlinear) equations as in the modified H-Q algorithm as described in the previous section.

In Figs. 4(a) and 4(b), a complex-valued object and its autocorrelation, respectively, are shown (only the left half of the latter is shown since it is Hermitian). For this example, Eqs. (5) were solved by Gaussian elimination using complex coefficients and the solution (the object) was found to be unique.

SOME UNIQUE CASES

Despite the phase-retrieval problem's not being unique, as demonstrated in Example 1 in the previous section, there are some specific classes of objects for which the solution is known to be unique. These unique objects have supports (or shapes) of special types.

Certain classes of objects having latent reference points can be reconstructed using a simpler recursive algorithm than the one described in the previous section. The simpler recursive algorithm[11] selects the order in which the equations are solved such that at each step one must solve only a single linear equation for a single unknown, which is a trivial computation that always gives a unique result. It is required that no division by zero be allowed, and this is ensured by the requirement that the values of the latent reference points not be zero. The latent reference points act in a similar manner to reference points for holography, but they do not initially satisfy the holographic separation condition.

1+0i	1+i	2-i	-3+2i	0-i
2-2i	2+2i	1-2i	2+i	-1+0i
1+3i	4+i	-2+3i	-1-i	1+2i
1-2i	2+2i	0+i	1-i	-1+i

(a)

-1-i	1-i	-3+0i	11-i	-2+0i
-3-2i	6-4i	-9-5i	9-9i	12-5i
-1-10i	2-i	-4+i	24-22i	9+11i
2+5i	1-23i	18+7i	-17-24i	114+0i
-2-5i	-4+9i	2-9i	6+7i	9-11i
-4+3i	0-14i	-3-5i	20+5i	12+5i
2+i	-9+6i	1-13i	4+i	-2+0i

(b)

Fig. 4. Example 3. A complex-valued object (a), which is uniquely related to its autocorrelation function (b). The right half of (b) (not shown) is the Hermitian conjugate of the left half (shown). For this example Eqs. (5) have a unique solution.

Examples of objects that can be uniquely reconstructed in this manner include (Fiddy *et al.*[16]) objects within a rectangle plus a point off one corner of the rectangle and objects having other supports as well.[11] In most cases the support of the object must be known *a priori* in order to ensure that one obtains a unique reconstruction, since it is usually not possible to deduce the support of the object from the support of its autocorrelation.[10] However, for the objects of Fiddy *et al.*[16] the support can be deduced from the autocorrelation support, and so the reconstruction in that case is unconditionally unique.[11] The objects may be complex valued. Furthermore, for these cases the boundary values need not be known *a priori* since they are computed in the first step of the recursive algorithm.[11,12]

CONCLUSIONS

Although boundary conditions are a powerful constraint for the phase-retrieval problem, it has been proved by counterexample (Example 1) that knowledge of the boundary conditions (the values of the edges of the object) is not sufficient to ensure a unique solution. In practice it may be that a unique solution is usually obtained simply because 2-D phase retrieval is usually unique even when the boundary conditions are not known.[14] It is not yet known what extra constraints are necessary to ensure uniqueness in general.

What seems to be more important to ensure uniqueness is that the object's support be a member of a special class of supports. It is not yet known in general exactly what properties the support must have (except for the special cases mentioned in the previous section) to ensure uniqueness; but it is known that objects with separated supports[17,18] are more likely to be unique (even in the one-dimensional case) and objects having complicated supports tend to be easier to reconstruct than objects with convex symmetric support in the 2-D case.[19]

Even when the underlying phase-retrieval problem is unique, it has been proved by counterexample (Example 2) that the equations of the H-Q recursive algorithm[12] [Eqs. (5)] does not necessarily have a unique solution. This is true even when the condition that $\alpha(m)$ not be proportional to $\beta(M-1-m)$ holds. This condition may be necessary to avoid ambiguities, but it is not sufficient. Clearly, for Eqs. (5) to be uniquely solvable it is necessary that the underlying phase-retrieval problem be unique, which is ensured if the object's Fourier transform is an irreducible polynomial, which is usually the case.[14] It is not now known what extra conditions are sufficient to ensure uniqueness for the solution of Eqs. (5), nor is it known whether it will usually be unique. Even if the solution to Eqs. (5) is not unique, one can find the solution or solutions if one uses the modified algorithm suggested here, which employs more of the points of the autocorrelation.

The value of the recursive algorithms may be more in their predictions of uniqueness than in their ability to reconstruct images, since they tend to be very sensitive to noise.[11,12] A more stable reconstruction method would be the iterative Fourier transform algorithm,[6] which repeatedly reinforces both the measured data and the *a priori* constraints on the reconstructed image.

ACKNOWLEDGMENT

This research was supported by the U. S. Air Force Office of Scientific Research under contract F49620-82-K-0018.

REFERENCES

1. R. N. Bracewell, *The Fourier Transform and Its Applications*, 2nd ed. (McGraw-Hill, New York, 1978).
2. C. Y. C. Liu and A. W. Lohmann, "High resolution image formation through the turbulent atmosphere," Opt. Commun. **8**, 372–377 (1973).
3. J. W. Goodman, "Analogy between holography and interferometric image formation," J. Opt. Soc. Am. **60**, 506–509 (1970).
4. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik (Stuttgart) **35**, 237–246 (1972); W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy* (Academic, New York, 1978); R. H. Boucher, "Convergence of algorithms for phase retrieval from two intensity distributions," Proc. Soc. Photo-Opt. Instrum. Eng. **231**, 130–141 (1980).
5. W. J. Dallas, "Digital computation of image complex amplitude from image- and diffraction-intensity: an alternative to holography," Optik **44**, 45–59 (1975).
6. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**, 2758–2769 (1982); "Reconstruction of an object from the modulus of its Fourier transform," Opt. Lett. **3**, 27–29 (1978); "Space object imaging through the turbulent atmosphere," Opt. Eng. **18**, 529–534 (1979).
7. P. J. Napier and R. H. T. Bates, "Inferring phase information from modulus information in two-dimensional aperture synthesis," Astron. Astrophys. Suppl. Ser. **15**, 427–430 (1974).
8. H. A. Arsenault and K. Chalasinka-Macukow, "The solution to the phase retrieval problem using the sampling theorem," Opt. Commun. **47**, 380–386 (1983).
9. G. H. Stout and L. H. Jensen, *X-Ray Structure Determination* (Macmillan, London, 1968).
10. J. R. Fienup, T. R. Crimmins, and W. Holsztynski, "Reconstruction of the support of an object from the support of its autocorrelation," J. Opt. Soc. Am. **72**, 610–624 (1982).
11. J. R. Fienup, "Reconstruction of objects having latent reference points," J. Opt. Soc. Am. **73**, 1421–1426 (1983).
12. M. H. Hayes and T. F. Quatieri, "Recursive phase retrieval using boundary conditions," J. Opt. Soc. Am. **73**, 1427–1433 (1983).
13. M. H. Hayes and T. F. Quatieri, "The importance of boundary conditions in the phase retrieval problem," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Institute of Electrical and Electronics Engineers, New York, 1982), pp. 1545–1548.
14. Yu. M. Bruck and L. G. Sodin, "On the ambiguity of the image reconstruction problem," Opt. Commun. **30**, 304–308 (1979).
15. J. R. Fienup, "Image reconstruction for stellar interferometry," in *Current Trends in Optics*, F. T. Arecchi and F. R. Aussenegg, eds. (Taylor and Francis, London, 1981), pp. 95–102.
16. M. A. Fiddy, B. J. Brames, and J. C. Dainty, "Enforcing irreducibility for phase retrieval in two dimensions," Opt. Lett. **8**, 96–98 (1983).
17. A. H. Greenaway, "Proposal for phase recovery from a single intensity distribution," Opt. Lett. **1**, 10–12 (1977); T. R. Crimmins and J. R. Fienup, "Ambiguity of phase retrieval for functions with disconnected supported," J. Opt. Soc. Am. **71**, 1026–1028 (1981).
18. T. R. Crimmins and J. R. Fienup, "Uniqueness of phase retrieval for functions with sufficiently disconnected support," J. Opt. Soc. Am. **73**, 218–221 (1983).
19. J. R. Fienup, "Experimental evidence of the uniqueness of phase retrieval from intensity data," in *Indirect Imaging*, J. A. Roberts, ed., Conference Proceedings of URSI/IAU Symposium, Sydney, Australia, August 30–September 2, 1983 (Cambridge U. Press, Cambridge, 1984), pp. 99–109.

Appendix G
AN ESTIMATOR OF FOURIER INTENSITIES IN STELLAR SPECKLE INTERFEROMTERY

C.C. Wackerman and J.R. Fienup

Appendix G

An Estimator of Fourier Intensities in Stellar Speckle Interferometry

C.C. Wackerman and J.R. Fienup

Environmental Research Institute of Michigan
P.O. Box 8618, Ann Arbor, Michigan 48107 USA

Abstract

For imaging objects through a turbulent atmosphere, speckle interferometry traditionally averages the Fourier intensities of many short-exposure degraded images to generate an estimate of the object's Fourier intensity. In this paper a model for the probability density of the short-exposure Fourier intensities is derived, from which an iterative procedure is developed which generates the maximum likelihood estimator of the object's Fourier intensity from the short exposure images.

AN ESTIMATOR OF FOURIER INTENSITIES IN STELLAR SPECKLE INTERFEROMETRY

1 INTRODUCTION

When imaging with a large earth-bound optical telescope, atmospheric turbulence degrades the resolution to approximately one second of arc, which is many times worse than the diffraction limit set by the telescope aperture. This situation can be improved by gathering Fourier intensity (power spectrum) data. Labeyrie [1] showed that if one takes a number of short-exposure images, one can combine the intensities (squared magnitudes) of their Fourier transforms to form an estimate of the object's Fourier transform intensity that has energy out to the diffraction limit of the telescope. This technique, known as stellar speckle interferometry, was demonstrated practically first by Gezari et al. [2] in 1972 and has been in use since. How to best combine the Fourier intensities of the short exposure images in order to get an estimate of the object's Fourier intensity is the subject of this paper. An iterative maximum likelihood estimator will be developed based upon a model for the probability density function of the Fourier intensity data that is calculated from the short-exposure images.

Let $o(\alpha,\beta)$ be the intensity distribution of the object we wish to image, and $h(\alpha,\beta)$ be a point spread function which includes both atmospheric and telescope effects. We have then [3, p. 259]

$$g(\alpha,\beta) = o(\alpha,\beta) * h(\alpha,\beta) \tag{1}$$

where $g(\alpha,\beta)$ is the image irradiance we record and * is the convolution operator. In this paper we are considering only the effects of atmospheric turbulence and not of noise. Letting upper case letters represent Fourier transforms of the corresponding lower case letters, then

$$G(u,v) = \iint g(\alpha,\beta)e^{-i2\pi(u\alpha+v\beta)}\, d\alpha\, d\beta$$

and, from Eq. (1),

$$|G(u,v)|^2 = |O(u,v)|^2\, |H(u,v)|^2 \qquad (2)$$

where H(u,v) is the optical transfer function for the imaging system which includes the effects of both the atmospheric turbulence and the telescope. Throughout this paper, the limits of integration are assumed to be $-\infty$ to $+\infty$ unless otherwise noted. Since we record a number of these Fourier intensities, each having the same contribution from the object and the telescope but a different contribution from the atmosphere, we have

$$|G_k(u,v)|^2 = |O(u,v)|^2\, |H_k(u,v)|^2,\ k = 1, \ldots, N \qquad (3)$$

where the subscript k is used to indicate quantities that change for each of the N short exposure images.

For a given Fourier plane coordinate, (u_0,v_0), the N real values that we have recorded, $|G_k(u_0,v_0)|^2$ for $k = 1, \ldots, N$, are the data values we need to combine in order to produce an estimate for the real parameter $|O(u_0,v_0)|^2$, the object's Fourier intensity (power spectrum), which we are trying to estimate.

We will generate an estimator by first deriving, in Section 2, the probability density function for the squared magnitude of the optical transfer function, $|H(u,v)|^2$, as a function of position in the frequency plane, the parameters of the optical system, and the atmospheric coherence diameter, r_0. In Section 3 we will then derive the density function of the observation $|G(u,v)|^2$ as a function of the parameter $|O(u,v)|^2$ and will describe an algorithm for generating the maximum likelihood estimator of $|O(u,v)|^2$ from the observations using this density function. Finally, Section 4 will have concluding comments.

2
PROBABILITY DENSITY FOR THE SQUARED MAGNITUDE OF THE OPTICAL TRANSFER FUNCTION

The optical transfer function, H(u,v), can be written as [4, p. 404]

$$H(u,v) = \frac{\iint P(x,y)P^*(x - \lambda fu, y - \lambda fv)\exp[X_1 + X_2 + i(\phi_1 - \phi_2)]\, dx\, dy}{\iint |P(x,y)|^2 \exp[2X]\, dx\, dy} \tag{4}$$

where λ = the wavelength of the propagating wavefront;

f = the focal length of the optical system;

$P(x,y)$ = the complex pupil function of the optical system;

$X_1 = X(x,y)$, the amplitude fluctuations caused by the atmosphere;

$X_2 = X(x - \lambda fu, y - \lambda fv)$;

$\phi_1 = \phi(x,y)$, the phase fluctuations, in radians, caused by the atmosphere;

$\phi_2 = \phi(x - \lambda fu, y - \lambda fv)$.

$X(x,y)$ and $\phi(x,y)$ are defined so that an on-axis plane wave with intensity I_0, after propagating through the atmosphere, will have the form $\sqrt{I_0}\exp[X(x,y) + i\phi(x,y)]$ at the entrance pupil of the optical system. Under the Rytov approximation to the solution of the wave equation [4, Sec. 8.4.3] (which assumes that the wavefront after propagating through a turbulent atmosphere can be modelled as a multiplicative perturbation of a wavefront that propagated through a constant, non-turbulent atmosphere) $X(x,y)$ and $\phi(x,y)$ are Gaussian random processes and $\phi(x,y)$ has a mean of zero [5, p. 6-18].

In order to investigate the variance of $X(x,y)$ and $\phi(x,y)$ consider the structure function of a random process, $A(x,y)$,

$$D_A(x,y,x',y') = E\left\{[A(x,y) - A(x',y')]^2\right\}, \tag{5}$$

where $E\{\ \}$ denotes expected value. If $A(x,y)$ is homogeneous, then Eq. (5) reduces to

$$D_A(\Delta x,\Delta y) = E\left\{[A(x,y) - A(x - \Delta x,\ y - \Delta y)]^2\right\} . \qquad (6)$$

The statistical properties of $X(x,y)$ and $\phi(x,y)$ are usually stated in terms of their structure functions, $D_X(\Delta x,\Delta y)$ and $D_\phi(\Delta x,\Delta y)$ (where it is assumed that the atmospheric turbulence is homogeneous). These in turn can be related back to the structure function of the refractive index of the atmosphere, $n(x,y,z)$, which has random fluctuations due mainly to random small-scale changes in temperature [4, Sec. 8.4.2]. Work by Kolmogorov [6] showed that the structure function for $n(x,y,z)$ can be represented by

$$D_n(\Delta x,\Delta y,\Delta z) = C_n^2(\Delta x^2 + \Delta y^2 + \Delta z^2)^{1/3} \qquad (7)$$

where C_n^2 is called the structure constant of the refractive index fluctuations and is a measure of the strength of those fluctuations [4, Sec. 8.4.2]. From these models, one can calculate the wave structure function, which is the sum of the structure functions for $X(x,y)$ and $\phi(x,y)$, as [4, p. 453]

$$D(\Delta x,\Delta y) = D_X(\Delta x,\Delta y) + D_\phi(\Delta x,\Delta y) = 6.884\left[\frac{[\Delta x^2 + \Delta y^2]^{1/2}}{r_o}\right]^{5/3} \qquad (8)$$

where r_o is the atmospheric coherence diameter defined by Fried [7].

We shall assume that variations in the amplitude of the distorted wavefront are so small that they can be ignored, so that $X(x,y) = 0$ and $D(\Delta x,\Delta y) = D_\phi(\Delta x,\Delta y)$. Equation (4) then becomes

$$H(u,v) = P_o^{-1} \iint P(x,y)\, P^*(x - \lambda f u,\ y - \lambda f v)$$

$$\exp\Big(i[\phi(x,y) - \phi(x - \lambda f u,\ y - \lambda f v)]\Big)\, dx\, dy \qquad (9)$$

where

$$P_o = \iint |P(x,y)|^2\, dx\, dy \qquad (10)$$

is a constant for each imaging system.

For a specific frequency location (u_o, v_o), the integration in Eq. (9) can be approximated as the sum of a large number of complex values. If we let $H(u_o, v_o) = R + iI$, and let Var[R], Var[I], E[R] and E[I] be the variance of the real and imaginary parts and the mean of the real and imaginary parts, respectively, of $H(u_o, v_o)$, then the probability density function of $|H(u_o, v_o)|$ (assuming that E[I] = 0, which is shown below) is [8, p. 124]

$$P_m(m) = \frac{m}{2\pi\sqrt{\text{Var[R]Var[I]}}} \int_0^{2\pi} \exp\left[-\frac{(m\cos\theta - E[R])^2}{2\,\text{Var[R]}} - \frac{m^2 \sin^2\theta}{2\,\text{Var[I]}}\right] d\theta \qquad (11)$$

where $m = |H(u_o, v_o)|$. We actually need the density function of $t = |H(u_o, v_o)|^2$ which, using Eq. (11), is [9, p. 90]

$$P_t(t) = \frac{1}{4\pi\sqrt{\text{Var[R]Var[I]}}} \int_0^{2\pi} \exp\left[-\frac{(\sqrt{t}\cos\theta - E[R])^2}{2\,\text{Var[R]}} - \frac{t \sin^2\theta}{2\,\text{Var[I]}}\right] d\theta. \qquad (12)$$

We will now generate expressions for the three parameters E[R], Var[R] and Var[I] which are needed to evaluate Eq. (12) as well as show that E[I] = 0.

From Eq. (9), assuming that P(x,y) is strictly real,

$$R = P_o^{-1} \iint P(r)P(r - \Delta) \cos\,[\phi(r) - \phi(r - \Delta)]\, dr \qquad (13)$$

where r is the vector (x,y), Δ is the vector $(\lambda f u_0, \lambda f v_0)$, and $dr = dxdy$. Likewise,

$$I = P_0^{-1} \iint P(r)P(r-\Delta) \sin [\phi(r) - \phi(r-\Delta)]\, dr. \tag{14}$$

From Eq. (13)

$$E[R] = P_0^{-1} \iint P(r)P(r-\Delta)\, E\left\{\cos [\phi(r) - \phi(r-\Delta)]\right\} dr. \tag{15}$$

As shown in Appendix A, for θ a zero-mean Gaussian random variable with variance σ^2, $E[\cos \theta] = \exp [-0.5\sigma^2]$. Thus Eq. (15) becomes

$$E[R] = P_0^{-1} \exp \left[-\tfrac{1}{2} D(\Delta)\right] \iint P(r)P(r-\Delta)\, dr \tag{16}$$

where

$$D(\Delta) = D(\lambda f u_0, \lambda f v_0) = 6.884 \left[\frac{\lambda f (u_0^2 + v_0^2)^{1/2}}{r_0}\right]^{5/3}. \tag{17}$$

Likewise, since $E[\sin \theta] = 0$ for θ zero-mean Gaussian (see Appendix A),

$$E[I] = 0. \tag{18}$$

Also from Eq. (9),

$$R^2 = P_0^{-2} \iiiint P(r)P(r-\Delta)P(r')P(r'-\Delta) \cos [\phi(r) - \phi(r-\Delta)] \cos [\phi(r') - \phi(r'-\Delta)]\, dr\, dr' \tag{19}$$

where r' is the vector (x',y') and $dr' = dx'\, dy'$. From Eq. (19) we have

$$E[R^2] = P_0^{-2} \iiiint P(r)P(r-\Delta)P(r')P(r'-\Delta)\, E\left\{\cos [\phi(r) - \phi(r-\Delta)] \cos [\phi(r') - \phi(r'-\Delta)]\right\} dr\, dr'. \tag{20}$$

The expectation in Eq. (20) can be rewritten as

$$E\left\{\cos\left[\phi(r) - \phi(r-\Delta)\right]\cos\left[\phi(r') - \phi(r'-\Delta)\right]\right\}$$

$$= \frac{1}{2} E\left\{\cos\left[\phi(r) - \phi(r-\Delta) + \phi(r') - \phi(r'-\Delta)\right]\right\}$$

$$+ \frac{1}{2} E\left\{\cos\left[\phi(r) - \phi(r-\Delta) - \phi(r') + \phi(r'-\Delta)\right]\right\} \qquad (21)$$

by using the trigometric identity cos A cos B = 1/2[cos(A + B) + cos(A - B)]. Each of the cosine arguments in Eq. (21) are zero mean Gaussian random variables (recall that $\phi(r)$ is zero mean). Thus, as mentioned above, we can evaluate the expectations if we know the variance of the cosine arguments. If we let $\phi_1 = \phi(r) - \phi(r-\Delta) + \phi(r') - \phi(r'-\Delta)$, then

$$\mathrm{Var}[\phi_1] = E[\phi_1^2] = E\left\{[\phi(r) - \phi(r-\Delta)]^2 + [\phi(r') - \phi(r'-\Delta)]^2 + 2[\phi(r) - \phi(r-\Delta)][\phi(r') - \phi(r'-\Delta)]\right\} . \qquad (22)$$

This can be rewritten as

$$\mathrm{Var}[\phi_1] = E\left\{[\phi(r) - \phi(r-\Delta)]^2\right\} + E\left\{[\phi(r') - \phi(r'-\Delta)]^2\right\}$$

$$+ E\left\{[\phi(r) - \phi(r'-\Delta)]^2 - [\phi(r) - \phi(r')]^2 - [\phi(r-\Delta) - \phi(r'-\Delta)]^2 + [\phi(r') - \phi(r-\Delta)]^2\right\} . \qquad (23)$$

Using the definition of the wave structure function in Eq. (6), Eq. (23) can be written as

$$\mathrm{Var}[\phi_1] = D(\Delta) + D(\Delta) + D(r - r' + \Delta) - D(r - r') - D(r - r') + D(r' - r + \Delta)$$

$$= 2[D(\Delta) - D(r - r') + \frac{1}{2} D(r - r' + \Delta) + \frac{1}{2} D(r' - r + \Delta)]. \qquad (24)$$

The first expectation on the right hand side of Eq. (21) is thus

$$E\left\{\cos\left[\phi(r) - \phi(r-\Delta) + \phi(r') - \phi(r'-\Delta)\right]\right\}$$
$$= \exp\left[-D(\Delta) + D(r-r') - \tfrac{1}{2}D(r-r'+\Delta) - \tfrac{1}{2}D(r'-r+\Delta)\right]. \quad (25)$$

Similarly, the second expectation on the left hand side of Eq. (21) can be evaluated as

$$E\left\{\cos\left[\phi(r) - \phi(r-\Delta) - \phi(r') + \phi(r'-\Delta)\right]\right\}$$
$$= \exp\left[-D(\Delta) - D(r-r') + \frac{1}{2}D(r-r'+\Delta) + \frac{1}{2}D(r'-r+\Delta)\right]. \quad (26)$$

Combining Eqs. (25) and (26) gives

$$E\left\{\cos\left[\phi(r) - \phi(r-\Delta)\right]\cos\left[\phi(r') - \phi(r'-\Delta)\right]\right\}$$
$$= \tfrac{1}{2}\exp\left[-D(\Delta)\right]\left(\exp\left[D(r-r') - \tfrac{1}{2}D(r-r'+\Delta) - \tfrac{1}{2}D(r'-r+\Delta)\right]\right.$$
$$\left. + \exp\left[-D(r-r') + \tfrac{1}{2}D(r-r'+\Delta) + \tfrac{1}{2}D(r'-r+\Delta)\right]\right). \quad (27)$$

Thus, Eq. (20) becomes

$$E[R^2] = \tfrac{1}{2}P_o^{-2}\exp\left[-D(\Delta)\right]\iiiint P(r)P(r-\Delta)P(r')P(r'-\Delta)$$
$$\left(\exp\left[D(r-r') - \tfrac{1}{2}D(r-r'+\Delta) - \tfrac{1}{2}D(r'-r+\Delta)\right]\right.$$
$$\left. + \exp\left[-D(r-r') + \tfrac{1}{2}D(r-r'+\Delta) + \tfrac{1}{2}D(r'-r+\Delta)\right]\right)\, dr\, dr'. \quad (28)$$

Note, however, that the wave structure functions in Eq. (28) depend only on the difference $r - r'$. We can thus do a change of variable, letting $p = r - r'$ and $q = r + r'$. The Jacobian of the transformation is then 1/4, so Eq. (28) becomes

$$E[R^2] = \frac{1}{8} P_o^{-2} \exp[-D(\Delta)] \iint L(p,\Delta)$$
$$\left(\exp[D(p) - \tfrac{1}{2} D(p+\Delta) - \tfrac{1}{2} D(-p+\Delta)]\right.$$
$$\left. + \exp[-D(p) + \tfrac{1}{2} D(p+\Delta) + \tfrac{1}{2} D(-p+\Delta)]\right) dp \quad (29)$$

where $L(p,\Delta)$ is the overlap integral

$$L(p,\Delta) = \iint P\left(\frac{q+p}{2}\right) P\left(\frac{q-p}{2}\right) P\left(\frac{q+p-2\Delta}{2}\right) P\left(\frac{q-p-2\Delta}{2}\right) dq. \quad (30)$$

Combining Eqs. (29) and (16) gives

$$Var[R] = E[R^2] - (E[R])^2$$
$$= \frac{1}{8} P_o^{-2} \exp[-D(\Delta)] \iint L(p,\Delta)$$
$$\left(\exp[D(p) - \tfrac{1}{2} D(p+\Delta) - \tfrac{1}{2} D(-p+\Delta)]\right.$$
$$\left. + \exp[-D(p) + \tfrac{1}{2} D(p+\Delta) + \tfrac{1}{2} D(-p+\Delta) - 2\right) dp. \quad (31)$$

Following a similar argument, we have from Eq. (14),

$$I^2 = P_o^{-2} \iiiint P(r)P(r-\Delta)P(r')P(r'-\Delta)$$
$$\sin[\phi(r) - \phi(r-\Delta)] \sin[\phi(r') - \phi(r'-\Delta)]\, dr\, dr'. \quad (32)$$

Using the trigonometric identity:

$$\sin[\phi(r) - \phi(r-\Delta)] \sin[\phi(r') - \phi(r'-\Delta)]$$
$$= \frac{1}{2}\left(\cos[\phi(r) - \phi(r-\Delta) - \phi(r') + \phi(r'-\Delta)]\right.$$
$$\left. - \cos[\phi(r) - \phi(r-\Delta) + \phi(r') - \phi(r'-\Delta)]\right) \quad (33)$$

and Eqs. (25) and (26) gives

$$E[I^2] = \frac{1}{2} P_o^{-2} \exp[-D(\Delta)] \iiiint P(r)P(r-\Delta)P(r')P(r'-\Delta)$$
$$\left(\exp[-D(r-r') + \frac{1}{2} D(r-r'+\Delta) + \frac{1}{2} D(r'-r+\Delta)]\right.$$
$$\left. - \exp[D(r-r') - \frac{1}{2} D(r-r'+\Delta) - \frac{1}{2} D(r'-r+\Delta)]\right) dr\, dr'. \quad (34)$$

Using the same change of variable as before gives

$$E[I^2] = \frac{1}{8} P_o^{-2} \exp[-D(\Delta)] \iint L(p,\Delta)$$
$$\left(\exp[-D(p) + \frac{1}{2} D(p+\Delta) + \frac{1}{2} D(-p+\Delta)]\right.$$
$$\left. - \exp[D(p) - \frac{1}{2} D(p+\Delta) - \frac{1}{2} D(-p+\Delta)]\right) dp \quad (35)$$

where $L(p,\Delta))$ is defined in Eq. (30). Since $E[I] = 0$, we have that $Var[I] = E[I^2]$. Note that the sum $E[R^2] + E[I^2] = E[|H|^2]$ agrees with the known results [4, Eq. (8.8-28)].

As mentioned above, we have assumed that $P(x,y)$ is real. If in addition we assume that $P(x,y)$ is a circular disk (i.e., $P(x,y) = 1$ for $(x^2 + y^2) \leq \rho^2$ and 0 elsewhere), then the overlap integral in Eq. (30) as well as the integral in Eq. (16) become simply the area in the intersection of the circles involved. Korff [10] has a series of expressions for Eq. (30) depending on the orientation of the circles and thus also for Eq. (16) since it is just a degenerate case. Under these assumptions, $E[R]$ can be analytically calculated from Eq. (16), and $Var[R]$ and $Var[I]$ can be calculated from a numeric integration of only the double integrals in Eqs (31) and (35). These three values are the parameters we need to specify the density function of interest in Eq. (12). In addition, since from Eq. (8) the structure function $D(\Delta)$ depends only upon the magnitude of Δ and not on its angle, from Eq. (9) it is seen that the point statistics of H depend only on the magnitude of Δ if the pupil function is circularly symmetric. Consequently, for a circular pupil, Eqs. (16),

(29) and (35) depend only on the magnitude of Δ, and so Eq. (12) needs to be evaluated only along a line through the origin of the frequency plane instead of throughout the whole plane. Figure 1 shows plots of $E[R]$, $E[R^2]$ and $E[I^2]$ calculated from Eqs. (16), (29) and (35) respectively, versus distance from the frequency plane origin for the case where the pupil function diameter was one-half meter and $r_0 = 0.2$ m. Figure 2 shows the same plots for a diameter of one meter and Figure 3 shows the same plots for a diameter of four meters.

For some locations in the frequency plane, Eq. (12) can be simplified. Assume that the vector Δ is large enough (i.e., that we are far enough away from the origin in the frequency plane) so that

$$\exp\left[-\frac{1}{2}D(\Delta)\right] \ll 1. \tag{36}$$

Then, from Eq. (16), $E[R] \ll 1$. Also, if Eq. (36) is true, then for reasonable values of the physical parameters ($r_0 \geq 10$ cm and telescope aperture ≤ 5 meters) we verified by numeric calculation that

$$\exp\left[D(p) - \frac{1}{2}D(p+\Delta) - \frac{1}{2}D(-p+\Delta)\right] \ll \exp\left[-D(p) + \frac{1}{2}D(p+\Delta) + \frac{1}{2}D(-p+\Delta)\right] \tag{37}$$

for almost all values of p and Δ. Using Eqs. (29) and (35), Eq. (37) implies that

$$E[R^2] = E[I^2] = \frac{1}{8}P_0^{-2}\iint L(p)\left(\exp\left[-D(p) - D(\Delta) + \frac{1}{2}D(p+\Delta) + \frac{1}{2}D(-p+\Delta)\right]\right)dp. \tag{38}$$

Using $E[R] = 0$ and $E[R^2] = E[I^2]$ in Eq. (12) gives

$$P_t(t) = \frac{1}{2\ \mathrm{Var}[R]}\exp\left[-\frac{t}{2\ \mathrm{Var}[R]}\right] \tag{39}$$

which is simply the exponential density. We can specify where such a simplification is possibly by setting some small value, δ, such that Eq. (36) will be assumed to hold if

$$\exp\left[-\frac{1}{2} D(\Delta)\right] \leq \delta. \tag{40}$$

Using the definition of $D(\Delta)$ in Eq. (8), Eq. (40) can be rewritten as

$$|\Delta| \equiv \left[(\lambda f u)^2 + (\lambda f v)^2\right]^{1/2} \geq r_o\left[\frac{-2\ln(\delta)}{6.884}\right]^{3/5} \tag{41}$$

which determines the cutoff frequency below which the simplification of Eq. (39) no longer applies.

3
MAXIMUM LIKELIHOOD ESTIMATOR OF FOURIER INTENSITIES

Maximum likelihood estimators (MLE's) are well known and often used in estimation problems [see, for example, 11, pp. 99-107; 9, pp. 290-298; 12, pp. 65-72]. Much of their popularity comes from the following two facts: (1) if the MLE is unbiased (i.e. the expected value of the MLE is equal to the value we are trying to estimate) then the MLE most likely has the smallest variance of any unbiased estimator [12, p. 68], and (2) as the number of observations grows the MLE always asymptotically approaches a minimum variance, unbiased estimator [9, p. 298; 12, p. 71]. MLE's can be described as follows. Assume we have N observations of some random variable, and we have the joint probability density function of our N observations as a function of a parameter, c, we are trying to estimate. If the actual observed values are substituted into the variables representing those observations in the joint density function, then the density function becomes only a function of c. The value of c which maximizes this one-dimensional function is the maximum likelihood estimator of c for this specific set of observations. If we can analytically differentiate the joint density function with respect to c, then we can derive an explicit expression for the MLE by setting this derivative to zero and solving for c. The resulting expression, which will be a function of the N observations, will be the MLE for c. If such a differentiation is not possible, but we are able to generate the value of the joint density function given the values of the N observations and given a value for c, an iterative search can be implemented to find the value of c which, for a given set of N observations, maximizes the joint density function. This algorithm would then take as input the N observed values and would output the MLE.

In Section 2, the density function for the magnitude squared of the optical transfer function at some specific location in the frequency plane, $|H(u_o,v_o)|^2$, was generated and expressions for the parameters needed to evaluate the density function were derived. As discussed in Section 1, the parameter we are interested in estimating is the object's Fourier intensity, $|O(u_o,v_o)|^2$, and the data we observe is $|G_k(u_o,v_o)|^2$. If we let $t = |H_k(u_o,v_o)|^2$ (to be consistent with Section 2), $c = |O(u_o,v_o)|^2$ (to be consistent with the previous paragraph) and $s = |G_k(u_o,v_o)|^2$, then from Eq. (2), i.e. $s = ct$, we have [13, Ch 6]

$$P_s(s) = \frac{1}{c} P_t\left(\frac{s}{c}\right) \tag{42}$$

where $P_s(s)$ is the probability density function of the observation, s. The density function P_t is defined in Eq. (12), where the three parameters needed to evaluate Eq. (12) (i.e., E[R], Var[R], Var[I]) are described in Eqs. (16), (31) and (35) respectively. If we assume that we have made N independent observations, then the joint density function is just the product of the N individual density functions [13, p. 167],

$$P_s(s_1, s_2, \ldots, s_N) = \left(\frac{1}{c}\right)^N P_t\left(\frac{s_1}{c}\right) P_t\left(\frac{s_2}{c}\right) \ldots P_t\left(\frac{s_N}{c}\right). \tag{43}$$

The general expression resulting from substituting Eqs. (16), (31) and (35) into Eq. (12), then substituting this result into Eq. (43), is not easily differentiable, and so an explicit expression for the MLE of $c = |O(u_o,v_o)|^2$ is not easily generated. Under the simplification of Eq. (36) however, Eq. (43) becomes the joint *density function of N independent, exponentially distributed random numbers* which is differentiable and generates the result

$$\hat{c} = \left[E\left\{|H(u_o,v_o)|^2\right\}\right]^{-1} \frac{1}{N} \sum_{k=1}^{N} |I_k(u_o,v_o)|^2 \tag{44}$$

where $\hat{c}$ is the MLE of $c = |O(u_o,v_o)|^2$. This is simply the traditional Labeyrie method where, in practice, the expectaton in Eq. (44), which is the transfer function of the speckle interferometry process, is estimated by observing a point-like star under, hopefully, an atmosphere having the same statistics as the atmosphere through which the images of the object were collected. Note also that the expectation in Eq. (44) can be evaluated as the sum of Eqs. (29) and (35) which simplifies to

$$\begin{aligned} E\left\{|H(u_o,v_o)|^2\right\} &= E\left\{R^2 + I^2\right\} \\ &= E[2R^2] \\ &= \frac{1}{4} P_o^{-2} \iint L(p) \quad \exp\left[-D(p) - D(\Delta) + \frac{1}{2} D(p + \Delta)\right. \\ &\quad \left. + \frac{1}{2} D(-p + \Delta)\right] dp. \end{aligned} \tag{45}$$

for Δ satisfying Eq. (41).

For locations in the frequency plane where the simplification of Eq. (36) is not possible [i.e., locations closer to the origin than the limit in Eq. (41)] the traditional method of estimation in Eq. (44) is no longer a MLE. For these locations, an iterative procedure as described above would have to be implemented to generate a MLE. Given the actual values of the N observations, $s_1, \ldots, s_N$, and given a value for the parameter, c, Eq. (12) can be used to evaluate $P_t(s_i/c)$ for $i = 1, \ldots, N$, and Eq. (43) can be used to evaluate the joint density function $P_s(s_1, \ldots, s_N)$. This allows an iterative search procedure to find the value of c which maximizes Eq. (43) whose output would thus be the maximum likelihood estimate of $|O(u_o,v_o)|^2$ given the observations. Note that this is a one-dimensional search, so any of the one-dimensional algorithms can be used, for example a Fibonacii or a golden section search [14, sec. 29.4].

4
CONCLUSIONS

In this paper we have derived a procedure for generating the maximum likelihood estimator of an object's Fourier intensity from observations of the images that have been degraded by the atmosphere; no noise effects were considered. It was shown that the traditional averaging methods used in speckle interferometry generated a maximum likelihood estimator only for higher spatial frequencies. For the lower spatial frequencies, a procedure for generating the maximum likelihood estimator was described which involves an iterative, one-dimensional search for a maximum value where the function to be searched was determined by the position in the frequency plane, the parameters of the optical system used, and the atmospheric coherence diameter. Whether a more efficient procedure exists for the lower frequencies remains to be seen. In addition, the improvement in the estimation error of this procedure over the traditional approach remains to be evaluated.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contributions to this work by J.N. Cederquist and T.R. Crimmins. This research was supported by the U.S. Air Force Office of Scientific Research under Contract F49620-82-K-0018.

REFERENCES

1. A. Labeyrie, "Attainment of Diffraction-Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images," Astron. Astrophys. 6, 85-87 (1970).
2. D.Y. Gezari, A. Labeyrie, and R.V. Stachnik, "Speckle Interferometry: Diffraction-Limited Measurements of Nine Stars with the 200-inch Telescope," Astrophys. J. Lett. 173, L1-L5 (1972).
3. J.C. Dainty, Laser Speckle and Related Phenomena, Chapter 7, 2nd Edition (Springer Verlag, 1983).
4. J.W. Goodman, Statistical Optics, (John Wiley and Sons, 1985).
5. W. Wolfe, and G. Zissis, eds., The Infrared Handbook, (IRIA, the Environmental Research Institute of Michigan, 1978) (Library of Congress No. 77-90780).
6. A. Kolmogorov, in Turbulence, Classic Papers on Statistical Theory S.K. Friedlander, L. Topper eds. (Wiley-Interscience, 1961).
7. D.L. Fried, "Optical Resolution Through a Randomly Inhomogeneous Medium for Very Long and Very Short Exposures," J. Opt. Soc. Am. 56, 1372-1379 (1966).
8. P. Beckman and A. Spizzichino, The Scattering of Electromagnetic Waves from Rough Surfaces, (Macmillan 1963).
9. P. Meyer, Introductory Probability and Statistical Applications, 2nd edition (Addison-Wesley, 1970).
10. D. Korff, "Analysis of a Method of Obtaining Near-Diffraction-Limited Information in the Presence of Atmospheric Turbulence," J. Opt. Soc. Am. 63, 971-980 (1973).
11. P. Bickel, and K. Doksum, Mathematical Statistics: Basic Ideas and Selected Topics (Holden-Day Inc., 1977).
12. H. Van Trees, Detection, Estimation and Modulation Theory, (John Wiley and Sons, 1968).
13. J. Davenport, Probability and Random Processes (McGraw-Hill Inc., 1970).
14. W. Gellert, H. Kustner, M. Hellwich, and H. Kastrer, eds., The UNR Concise Encyclopedia of Mathematics, (Van Nostrand Reinhold, 1975).

APPENDIX A

Let ϕ be a Gaussian random variable with zero mean and variance σ^2. Then

$$E[\cos \phi] = \int_{-\infty}^{\infty} \cos \phi \, \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\phi^2}{2\sigma^2}\right] d\phi$$

$$= \sqrt{\frac{2}{\pi\sigma^2}} \int_{0}^{\infty} \cos(\phi) \exp\left[\frac{-\phi^2}{2\sigma^2}\right] d\phi$$

$$= \sqrt{\frac{2}{\pi\sigma^2}} \sqrt{\frac{\pi\sigma^2}{2}} \exp\left[-\frac{1}{2}\sigma^2\right]$$

$$= \exp\left[-\frac{1}{2}\sigma^2\right].$$

Similarly,

$$E[\sin \phi] = \int_{-\infty}^{\infty} \sin \phi \, \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-\phi^2}{2\sigma^2}\right] d\phi$$

$$= 0$$

since the sine is an odd function.

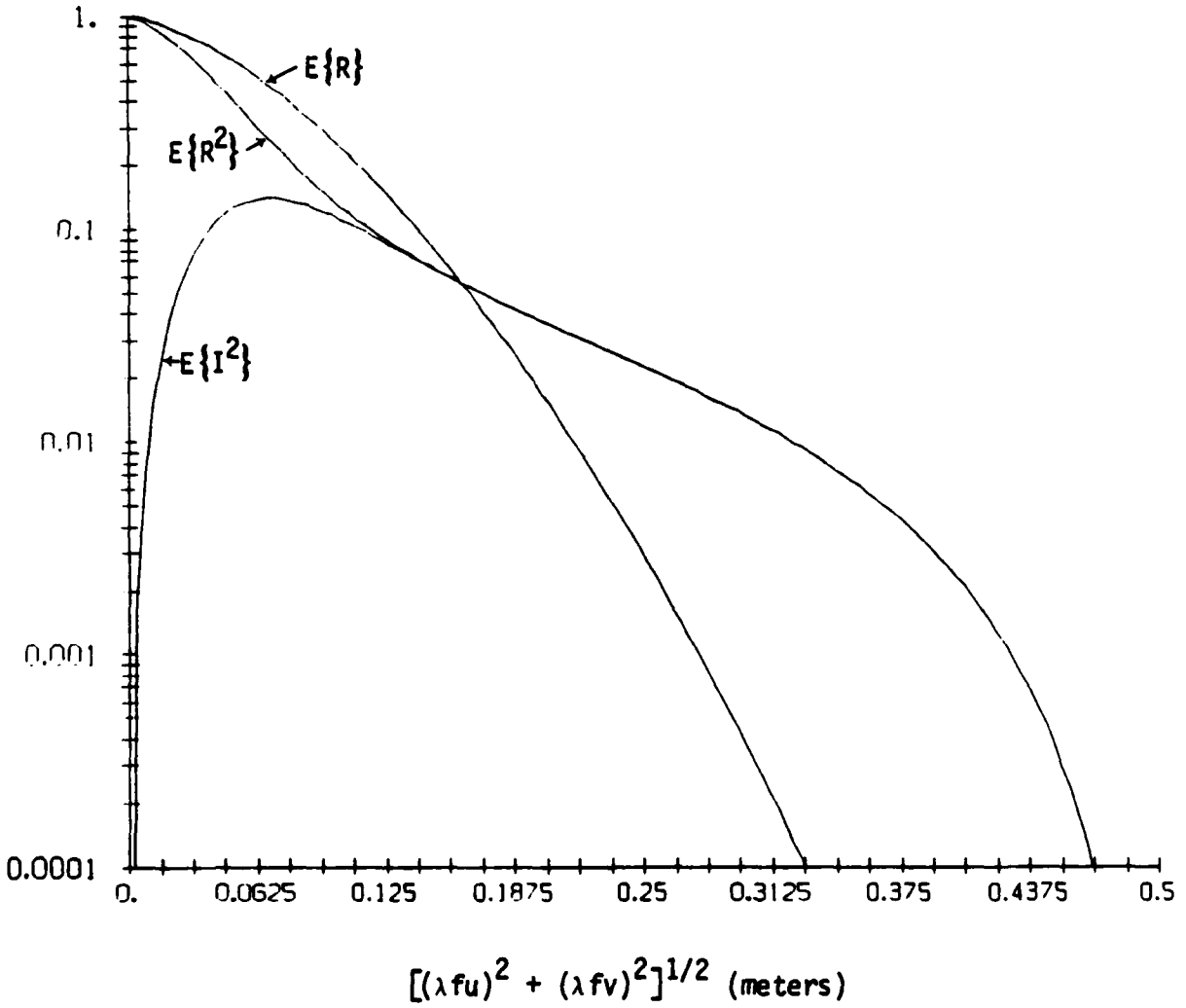


Figure 1. Expected Value of R, R^2 and I^2 for r_0 = 0.2 m, pupil diameter = .5 m.

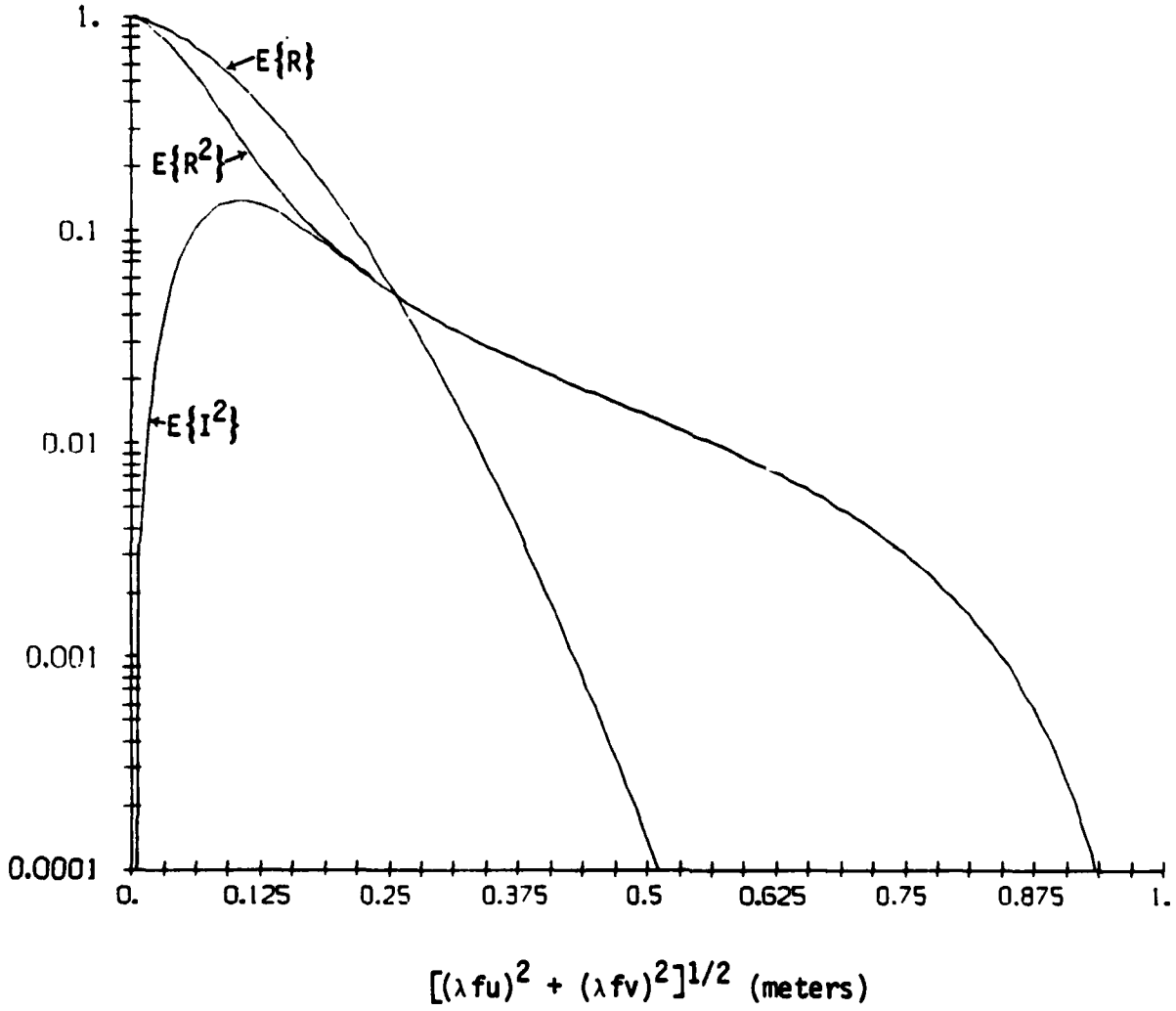


Figure 2. Expected Value of R, R^2 and I^2 for r_0 = 0.2 m, pupil diameter = 1 m.

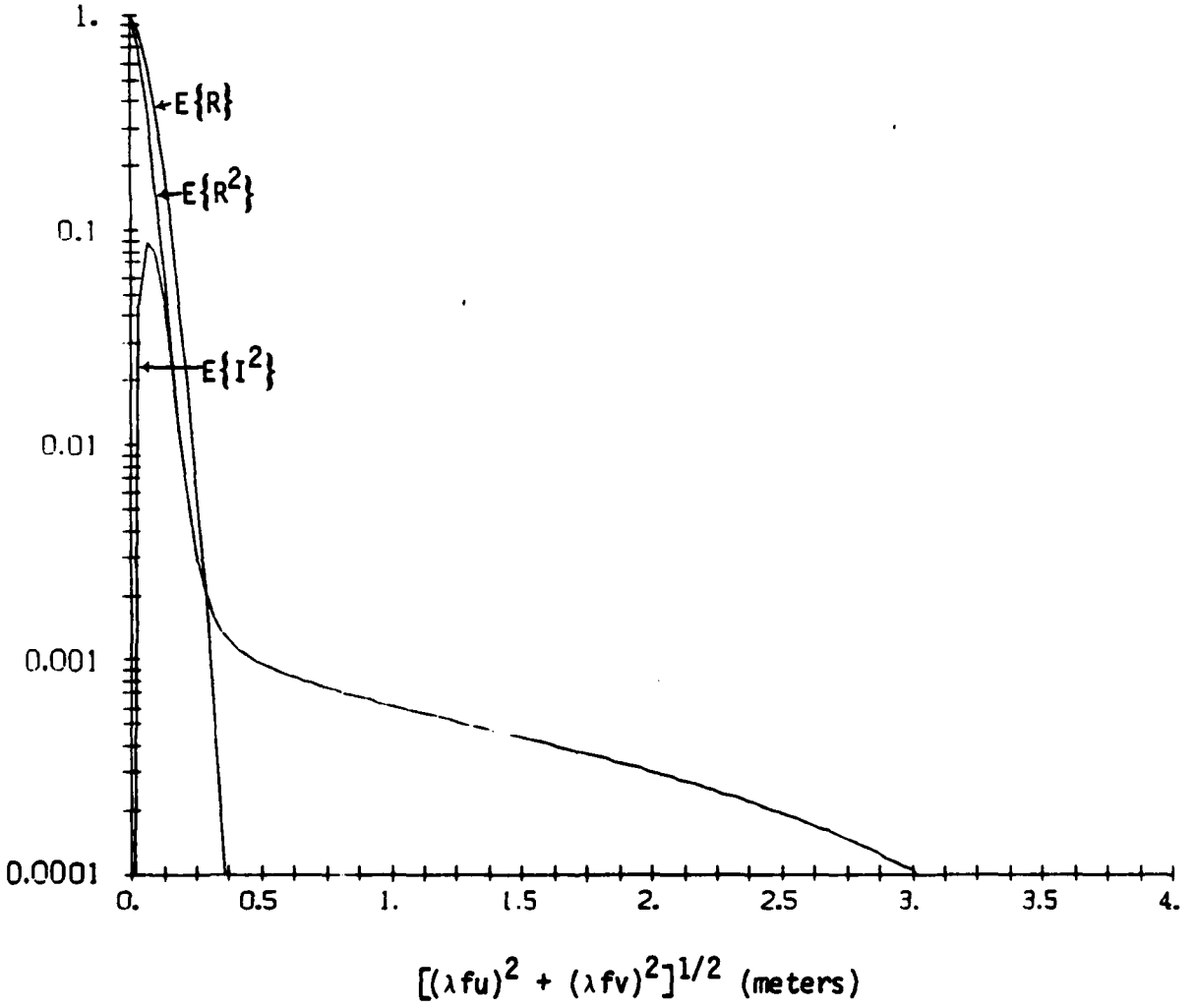


Figure 3. Expected Value of R, R^2 and I^2 for r_0 = 0.2 m, pupil diameter = 4m.

REFERENCES

1. A.A. Michelson and F.G. Pease, "Measurement of the Diameter of Alpha Orionis with the Interferometer," Astrophys. J. 53, 249-259 (1921).
2. R. Hanbury Brown and R.Q. Twiss, "Correlation Between Photons in Two Coherent Beams of Light," Nature 177, 27-29 (1956).
3. D.G. Currie, S.L. Knapp, and K.M. Liewer, "Four Stellar-Diameter Measurements by a New Technique: Amplitude Interferometry," Astrophys. J. 187, 131-144 (1974).
4. A. Labeyrie, "Attainment of Diffraction Limited Resolution in Large Telescopes by Fourier Analysing Speckle Patterns in Star Images," Astron. and Astrophys. 6, 85-87 (1970).
5. D.Y. Gezari, A. Labeyrie, and R.V. Stachnik, "Speckle Interferometry: Diffraction-Limited Measurements of Nine Stars with the 200-Inch Telescope," Astrophys. J. Lett. 173, L1-L5 (1972).
6. J.R. Fienup, "Reconstruction of an Object from the Modulus of Its Fourier Transform," Opt. Lett. 3, 27-29 (1978).
7. J.R. Fienup, "Space Object Imaging through the Turbulent Atmosphere," Opt. Eng. 18, 529-534 (1979).
8. J.R. Fienup, "Phase Retrieval Algorithms: A Comparison," Appl. Opt. 21, 2758-2769 (1982).
9. J.C. Dainty and J.R. Fienup, "Phase Retrieval and Image Reconstruction for Astronomy," Chapter 7 in H. Stark, ed., Image Recovery: Theory and Application (Academinc Press, 1986).
10. J.R. Fienup, "Diffraction-Limited Imaging of Space Objects II," Interim Scientific Report to AFOSR, ERIM Rept. No. 161900-13-T (October 1984).
11. J.R. Fienup, "Reconstruction and Synthesis Applications of an Iterative Algorithm," in Transformations in Optical Signal Processing, W.T. Rhodes, J.R. Fienup and B.E.A. Saleh, eds., Proc. SPIE 373, 147-160 (1981).
12. J.R. Fienup, "Diffraction-Limited Imaging of Space Objects I," Interim Scientific Report to AFOSR, ERIM Rept. No. 161900-6-T (May 1983).
13. J.R. Fienup, "Reconstruction of Objects Having Latent Reference Points," J. Opt. Soc. Am. 73, 1421-1426 (1983).
14. J.R. Fienup and C.C. Wackerman, "Phase Retrieval Stagnation Problems and Solutions," to appear in J. Opt. Soc. Am. A (November 1986).

15. J.R. Fienup and C.C. Wackerman, "Uniqueness of Phase Retrieval Shown Empirically," submitted to J. Opt. Soc. Am. A.

16. J.R. Fienup, "Reconstruction of a Complex-Valued Object from the Modulus of Its Fourier Transform Using a Support Constraint," to appear in J. Opt. Soc. Am. A (January 1987).

17. J.R. Fienup, "Phase Retrieval Using Boundary Conditions," J. Opt. Soc. Am. A. 3, 284-288 (1986).

18. C.C. Wackerman and J.R. Fienup, "An Estimator of Fourier Intensities in Stellar Speckle Interferometry," submitted to J. Opt. Soc. Am. A.

19. T.R. Crimmins and J.R. Fienup, "Uniqueness of Phase Retrieval for Functions with Sufficiently Disconnected Support," J. Opt. Soc. Am. 73, 218-221 (1983).

20. J.R. Fienup, "Phase Retrieval in Astronomy," in Signal Recovery and Synthesis with Incomplete Information and Partial Constraints, digest of papers (Opt. Soc. Am., 1983), Th A8, Jan. 12-14, 1983, Incline Village, NV.

21. J.R. Fienup, "Two-Dimensional Image Reconstruction Algorithms," in Information Processing in Astronomy and Optics, digest of papers (Opt. Soc. Am., 1983), Th A11, June 23-24, 1983, St. Paul, MN.

22. J.R. Fienup, "Holographic Reconstruction with Latent Reference Points," presented at the Annual Meeting of the Optical Society of America, New Orleans, LA, October 1983; Abstract: J. Opt. Soc. Am. 73, 1861 (1983).

23. J.R. Fienup, "Experimental Evidence of the Uniqueness of Phase Retrieval from Intensity Data," in Indirect Imaging, Proc. URSI/IAU Symposium, 30 Aug.-1 Sept., 1983, Sydney, Australia (Cambridge University Press, 1984), pp. 99-109.

24. J.R. Fienup, "Comments on 'The Reconstruction of a Sequence from the Phase or Magnitude of Its Fourier Transform'," IEEE Trans. Acoustics, Speech, Signal Processing ASSP-31, 738-739 (1983).

25. J.R. Fienup, "Comparison of Phase Retrieval Algorithms," Chapter 4 in T.S. Huang, ed., Advances in Computer Vision and Image Processing, Vol. 1, Image Reconstrucion from Incomplete Information (JAI Press, Greenwich, Conn., 1984), pp. 191-225.

26. J.R. Fienup, "Phase Retrieval from a Single Intensity Distribution," in Optics in Modern Science and Technology, Conference Digest for ICO-13, 20-24 August 1984, Sapporo, Japan, pp. 606-609.

27. J.R. Fienup, "Phase Retrieval and Image Reconstruction in Astronomy," ARO Workshop on Unconventional Imagery, Rigi-Kaltbad-First, Switzerland, 23-28 September 1984.

28. J.R. Fienup and C.C. Wackerman, "Improved Phase Retrieval Algorithm," presented at 1984 Annual Meeting of the Optical Society of America; Abstract: J. Opt. Soc. Am. A 1, 1320 (1984).

29. J.R. Fienup, "Phase Retrieval Using a Support Constraint," presented at IEEE ASSP Workshop on Multidimensional Digital Signal Processing, 28-30 October 1985, Leesburg, VA.

30. J.R. Fienup, "Phase Retrieval: Algorithm Improvements, Uniqueness and Complex Objects," O.S.A. topical meeting on Signal Recovery and Synthesis II, Technical Digest, 2-4 April 1986, Honolulu, Hawaii, pp. 40-43.

31. T.R. Crimmins and J.R. Fienup, "Phase Retrieval for Functions with Sufficiently Disconnected Support," in Signal Recovery and Synthesis with Incomplete Information and Partial Constraints, digest of papers to the OSA topical meeting, 12-14 January 1983, Incline Village, Nev.

32. J.R. Fienup, "Fine Resolution Imaging of Space Objects," Final Scientific Report to AFOSR, Contract No. F49620-80-C-0006, ERIM Report No. 145400-14-F, February 1982.

33. R.H.T. Bates and W.R. Fright, "Composite Two-Dimensional Phase-Restoration Procedure," J. Opt. Soc. Am. 73, 358-365 (1983).

34. J.R. Fienup, "Fourier Modulus Image Construction," Report RADC-TR-81-63 (1981).

35. M.H. Hayes and T.F. Quatieri, "Recursive Phase Retrieval Using Boundary Conditions," J. Opt. Soc. Am. 73, 1427-1433 (1983).

36. R.H.T. Bates and F.M. Cady, "Towards True Imaging by Wideband Speckle Interferometry," Opt. Commun. 32, 365-369 (1980).

37. F.M. Cady and R.H.T. Bates, "Speckle Processing Gives Diffraction-Limited True Images from Severely Aberrated Instruments," Opt. Lett. 5, 438-440 (1980).

38. C. Papaliolios, P. Nisenson and S. Ebstein, "Speckle Imaging with the PAPA Detector," Appl. Opt. 24, 287-292 (1985).

END

1-87

DTIC